의학박사 학위논문

Genome-wide association and expression quantitative trait loci studies for thyroid cancer and nodule

갑상선암과 결절에 대한 전장 유전체 연관 및 발현 양적 형질 유전자좌 연구

2020 년 2 월

서울대학교 대학원

의학과 중개의학 전공

황 보 율

Abstract

Genome-wide association and expression quantitative trait loci studies for thyroid cancer and nodule

Yul Hwangbo

Major in Translational Medicine

Department of Medicine

Seoul National University Graduate School

Thyroid cancer is the most common endocrine cancer and thyroid nodule is most common endocrine problem in Korea. Both phenotypes show a high degree of heritability. Several genome-wide association studies (GWAS) for thyroid cancer were conducted in European descendants and identified susceptibility loci of differentiated thyroid cancer (DTC). However, there is

no GWAS for thyroid cancer in Asian population, and inherited genetic risk factors for thyroid nodules and their associations with thyroid cancer remain unknown.

Here, GWAS and replication study was performed using a total of 1,085 DTC cases and 8,884 controls of Koreans and these results were validated with an expression quantitative trait loci (eQTL) analysis and clinical phenotypes. The most robust associations were observed in the *NRG1* gene (rs6996585, P=1.08×10^{-10}), and this SNP was also associated with *NRG1* expression in thyroid tissues. In addition, three previously reported loci (*FOXE1, NKX2-1,* and *DIRC3*) were confirmed and seven susceptibility loci (*VAV3, PCNXL2, INSR, MRSB3, FHIT, SEPT11,* and *SLC24A6*) associated with DTC were newly identified. Furthermore, I identified specific variants of DTC that have different effects according to the cancer type or ethnicity.

Furthermore, a three-stage GWAS for thyroid nodules was performed. The discovery stage involved a genome-wide scan of 811 subjects with thyroid nodules and 691 subjects with a normal thyroid from a population-based cohort. Replication studies were conducted in an additional 1981 cases and 3100 controls from the participants of a health check-up. Expression quantitative trait loci (eQTL) analysis was also performed using public data. The most robust

association was observed in *TRPM3* (rs4745021) in the joint analysis (OR=1.26, $P = 6.12 \times 10^{-8}$) and meta-analysis (OR = 1.28, $P = 2.11 \times 10^{-8}$). Signals at *MBIP/NKX2-1* were replicated but did not reach genome-wide significance in the joint analysis (rs2415317; $P = 4.62 \times 10^{-5}$, rs944289; $P = 8.68 \times 10^{-5}$). The eQTL analysis showed that *TRPM3* expression was associated with the rs4745021 genotype in thyroid tissues.

The results of GWAS for DTC provide deeper insight into the genetic contribution to thyroid cancer in different populations. And GWAS for thyroid nodule suggest that thyroid nodules have a genetic predisposition distinct from that of thyroid cancer.

· This doctoral dissertation is based on the following published research papers.

· Genome-wide association and expression quantitative trait loci studies identify multiple susceptibility loci for thyroid cancer. Nat Commun. 2017 Jul 13;8:15966.

· Genome-Wide Association Study Reveals Distinct Genetic Susceptibility of Thyroid Nodules From Thyroid Cancer. J Clin Endocrinol Metab. 2018 Dec 1;103(12):4384-4394.

· Genome-Wide Association Studies of Autoimmune Thyroid Diseases, Thyroid Function, and Thyroid Cancer. Endocrinol Metab (Seoul). 2018 Jun; 33(2): 175–184.

Keywords: Thyroid cancer, Thyroid nodule, Genome-Wide Association study, Expression quantitative trait loci, Genetic susceptibility

***Student Number*:** 2015-30903

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

PTC, papillary thyroid cancer

FTC, follicular thyroid cancer

MTC, medullary thyroid cancer

PDTC, poorly differentiated thyroid cancer

ATC, anaplastic thyroid cancer

DTC, differentiated thyroid cancer

GWAS, genome-wide association study

eQTL, expression quantitative trait loci

KCCR, Korea Central Cancer Registry

SNP, single nucleotide polymorphism

KoGES, Korean Genome and Epidemiology Study

KARE, Korean Association Resource

ETE, extrathyroidal extension

LN, lymph node

Q-Q, quantile-quantile

MAF, minor allele frequency

Chr, chromosome number

GSEA, Gene Set Enrichment Analysis

FOXE1, forkhead box E1

TTF, thyroid transcription factor

MBIP, MAP3K12 binding inhibitory protein 1

NKX2-1, NK2 homeobox 1

DIRC3, disrupted in renal carcinoma 3

NRG1, neuregulin 1

IMMP2L, inner mitochondrial membrane peptidase subunit 2

RARRES1, retinoic acid receptor responder 1

SNAPC4, small nuclear RNA activating complex polypeptide 4

BATF, basic leucine zipper ATF-like transcription factor

AP-1, activator protein 1

ATF, activating transcription factor

DHX35, DEAH-box helicase 35

ARSB, arylsulfatase B; *SPATA13*, spermatogenesis associated 13

GALNTL4,

UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase-like 4

FOXA2, forkhead box A2

WDR11-AS1, WDR11 antisense RNA 1

HTR1B, 5-hydroxytryptamine receptor 1B

PCNXL2, pecanex-like 2

OBFC1, oligosaccharide-binding folds containing 1

NREP, neuronal regeneration related protein

EPB41L4A, erythrocyte membrane protein band 4.1 like 4A

SMAD3, SMAD family member 3

TERC, telomerase RNA component

LRRC34, leucine rich repeat containing 34

TERT, telomerase reverse transcriptase

MSRB3, methionine sulfoxide reductase B3;

VAV3, vav guanine nucleotide exchange factor 3

SEPT11, septin 11; *FHIT*, fragile histidine triad

INSR, insulin receptor

SLC24A6, solute carrier family 24 member A6.

Introduction

1. Epidemiology of thyroid cancer

Thyroid cancer is the most common endocrine malignancy worldwide(1). According to the statistics of the Korea Central Cancer Registry (KCCR), 26,051 Koreans were newly diagnosed from thyroid cancer in 2016 (2). The estimated age-standardized incidence rate of thyroid cancer in Korea was 60.7 per 100,000 people, which was the highest in the world according to the The Global Cancer Observatory 2018 database of the International Agency for Research on Cancer/World Health Organization (http://gco.iarc.fr/). Major types of thyroid cancer include papillary thyroid cancer (PTC), follicular thyroid cancer (FTC), medullary thyroid cancer (MTC), poorly differentiated thyroid cancer (PDTC) and anaplastic thyroid cancer (ATC). PTC and FTC are the most predominant types of thyroid cancer, representing 96.5 % and 1.9 %, respectively, in Korea. PTC and FTC are often referred to as differentiated thyroid cancer (DTC).

2. Risk factors of differentiated thyroid cancer

The most important risk factor for thyroid cancer is female. DTC occurs three to four times more often in women than in men, and the reason is unclear (3).

Thyroid cancer can occur in all age groups, but it is relatively rare in childhood, and the incidence increases gradually with age. The risk of DTC peaks at ages 50 to 60 years (2).

Radiation exposure is the best-established risk factor for DTC (1). Sources of exposure include radiation fallout from nuclear accident and, therapeutic or diagnostic radiations to the head and neck. Exposure to high levels of radiation from nuclear accident at a young age is risk factor for DTC lasting for several decades (4). Head or neck irradiation for therapeutic purposes in childhood is a risk factor for thyroid cancer and the risk depends on dose of radiation and the age of the child. The low levels of radiation from medical imaging procedures such as x-rays and CT scans also could be associated with development of thyroid cancer (1). However, there is a lack of research on whether radiation exposure from imaging tests increases thyroid cancer.

Overweight and obesity are recently considered as risk factors for thyroid cancer (5). Several lines of evidence showed that high level of body mass index (BMI) is positively associated with risk for thyroid cancer and an increase in obese population may have contributed to the rise in thyroid cancer incidence (6, 7).

3. Heritability of differentiated thyroid cancer

Thyroid cancers show a high degree of heritability, with genetic factors accounting for more than 50 % of the causes of thyroid cancer. Except for MTC, which is well known to be caused by germline or somatic mutations, prevalence of familial DTC are reported in the range of 2.5%-11.3% (8-12). In western countries, the prevalence of familial DTC accounts for approximately 4-5% in patients with thyroid cancers of follicular cell origin (13, 14). In Korea, the prevalence of Familial DTC was also high and accounts for 9.6% (15), suggesting that the genetic susceptibility of DTC could differ from the western countries.

4. Familial syndromes associated with thyroid cancer and germline mutation of differentiated thyroid cancer

Only 5% of cases of familial DTC were reported to be of the syndromic form, which is accompanied by well-known germline mutations, including Cowden syndrome, familial adenomatous polyposis, Gardner syndrome, Carney complex type 1, Werner syndrome, and DICER1 syndrome (16). Recent next-generation sequencing study showed a germline variant in *HABP2* is associated with risk of familial DTC (17). Except for familial syndromes associated with

DTC and rare germline mutations of DTC, the majority of cases of familial DTC were found not to be caused by germline mutations, despite its pattern of genetic inheritance.

5. Epidemiology of thyroid nodule

Thyroid nodules are a common endocrine problem and their prevalence in the adult population is 19–68% as evaluated by high resolution ultrasonography (18). The prevalence of thyroid nodules especially increases with age, and the process of thyroid nodule development has been explained by an increased exposure to environmental factors or as an aging phenomenon (19, 20). However, the cause of benign thyroid nodules is unclear.

6. Clinical significance and heritability of thyroid nodule

The clinical importance of thyroid nodules rests with the possibility of thyroid cancer, which accounts for 7–15% of thyroid nodules (18). Similar to thyroid cancer, old age, female gender, obesity, smoking, iodine deficiency, and radiation exposure are known risk factors for thyroid nodules (21-23), and heritability also contributes to their genesis (24). Several evidences suggest that a family history of thyroid nodules is associated with an increased risk of

thyroid cancer (25, 26). Thus, the genetic risks of thyroid cancer and thyroid nodules could be closely related, and identification of the shared susceptible genetic loci might be useful to establish a management strategy for thyroid nodules, especially for patients with a family history.

7. Genome-wide association study for differentiated thyroid cancer

Previous researches into the genes responsible for thyroid disease has identified several candidates (27). However, candidate gene studies have been controversial and have shown very few reproducible findings. In the last decade, genome-wide association studies (GWAS) have been extensively used to identify genes involved in complex diseases (28). GWAS have facilitated the screening of a large proportion of the genome and discovered a variety of susceptibility genes. GWAS have been widely applied in autoimmune thyroid diseases, thyroid function, and thyroid cancer, and have identified susceptibility genes for thyroid-related phenotypes. The first GWAS of thyroid cancer was reported in 2009 and showed that common variants located on 9q22.33 (*FOXE1*) and 14q13.3 (*NKX2-1*) were associated with DTC (29). Associations at *FOXE1*, *MBIP/NKX2-1*, *DIRC3*, and *NRG1* have been identified and repeatedly confirmed in individuals of European ancestry (29-32). Several markers

associated with DTC, including *IMMP2L*, *RARRES1*, *SNAPC4*, *BATF*, *DHX35*, *GALNTL4*, *HTR1B*, *FOXA2*, and *WDR11-AS1*, were identified but not replicated in other studies (31-34). A recent meta-analysis of GWAS including a total of 3,001 DTC patients and 287,550 controls from 5 study groups of European populations found 5 novel loci (*PCNXL2*, *TERC*, *NREP- EPB41L4A*, *OBFC1*, and *SMAD3*) (35).

Table 1 provides the susceptibility loci identified in GWAS of thyroid cancer. The most robust signals were detected on 9q22.33 (*FOXE1*) in Caucasians (29, 35). The *FOXE1* locus was also reported to be a susceptibility gene for radiation-related thyroid cancer (36). A functional study showed that common variants on *FOXE1* regulated *FOXE1* transcription through the recruitment of the USF1/USF2 transcription factors (37). Several reports demonstrated that variants of *FOXE1* were related to aspects of the clinical aggressiveness of papillary thyroid cancer, such as tumor stage, size, lymphocytic infiltration, and extrathyroidal extension (38, 39).

8. Genetic studies for thyroid nodule

Regarding genetics of thyroid nodule, two studies of single nucleotide

polymorphism (SNP) analysis have reported an association between benign thyroid nodules and a SNP on *MBIP/NKX2-1*, which is a known DTC susceptibility locus (40, 41). These data suggest a common genetic etiology between benign thyroid nodules and DTC;

Table 1. DTC or PTC associated SNPs of previous European GWAS and replicated SNPs in Asia

Locus	Gene	SNP	OR	*P*-value	Population	Method	Ref.
9q22.33	*FOXE1*	rs965513	1.75	1.7×10^{-27}	Iceland etc.	GWAS	(29)
			1.65	4.8×10^{-12}	Belarus	GWAS	(36)
			1.69	1.3×10^{-4}	Japanese	SNP	(42)
			1.53	1.4×10^{-4}	Chinese	SNP	(40)
			1.59	4.2×10^{-4}	Japanese	SNP	(41)
		rs7028661	1.64	1.0×10^{-22}	Spain	GWAS	(34)
		rs10122541	1.54	1.1×10^{-17}	Spain	GWAS	(34)
		rs7037324	1.54	1.2×10^{-17}	Spain	GWAS	(34)
14q13.3	*NKX2-1*	rs944289	1.37	2.0×10^{-9}	Iceland etc.	GWAS	(29)
			1.24	1.5×10^{-5}	Spain	GWAS	(34)
			1.21	0.0121	Japanese	SNP	(42)
			1.53	2.2×10^{-10}	Chinese	SNP	(40)
			1.23	0.003	Japanese	SNP	(41)
		rs116909374	2.09	4.6×10^{-11}	Iceland etc.	GWAS	(30)
2q35	*DIRC3*	rs966423	1.34	1.3×10^{-9}	Iceland etc.	GWAS	(30)
			1.31	0.0010	Chinese	SNP	(40)
		rs6759952	1.21	6.4×10^{-10}	Italy etc.	GWAS	(31)
8p12	*NRG1*	rs2439302	1.36	2.0×10^{-9}	Iceland etc.	GWAS	(30)
			1.41	2.78×10^{-5}	Chinese	SNP	(40)
			1.27	0.003	Japanese	SNP	(41)
7q31.1	*IMMP2L*	rs10238549	1.27	4.1×10^{-6}	Italy etc.	GWAS	(31)
		rs7800391	1.25	5.7×10^{-6}	Italy etc.	GWAS	(31)
3q25.32	*RARRES1*	rs7617304	1.25	4.6×10^{-5}	Italy etc.	GWAS	(31)
9q34	*SNAPC4*	rs10781500	1.23	3.5×10^{-5}	Italy etc.	GWAS	(31)
14q24.3	*BATF*	rs10136427	1.40	4.4×10^{-7}	Italy etc.	GWAS	(32)
20q11.23	*DHX35*	rs7267944	1.39	2.1×10^{-8}	Italy etc.	GWAS	(32)
5q14	*ARSB*	rs13184587	1.28	8.5×10^{-6}	Italy etc.	GWAS	(32)
13q12	*SPATA13*	rs1220597	1.26	3.3×10^{-6}	Italy etc.	GWAS	(32)
11p15.3	*GALNTL4*	rs7935113	1.36	7.4×10^{-7}	Italy etc.	GWAS	(33)
20p11	*FOXA2*	rs1203952	1.29	4.4×10^{-6}	Italy etc.	GWAS	(33)
10q26.12	*WDR11-AS1*	rs2997312	1.35	1.2×10^{-4}	Spain	GWAS	(34)
		rs10788123	1.26	5.2×10^{-4}	Spain	GWAS	(34)
		rs1254167	1.38	5.9×10^{-5}	Spain	GWAS	(34)
6q14.1	*HTR1B*	rs4075570	0.82	2.0×10^{-4}	Spain	GWAS	(34)
1q42.2	*PCNXL2*	rs12129938	1.32	4.0×10^{-11}	Iceland etc	GWAS	(35)
3q26.2	*TERC-LRRC34*	rs6793295	1.23	2.7×10^{-8}	Iceland etc	GWAS	(35)
5p15.33	*TERT*	rs10069690	1.2	3.2×10^{-7}	Iceland etc	GWAS	(35)
5q22.1	*NREP-EPB41L4A*	rs73227498	1.37	3.0×10^{-10}	Iceland etc	GWAS	(35)
10q24.33	*OBFC1*	rs7902587	1.41	5.4×10^{-11}	Iceland etc	GWAS	(35)
15q22.33	*SMAD3*	rs2289261	1.23	3.1×10^{-9}	Iceland etc	GWAS	(35)

OR, odd ratio; SNP, single nucleotide polymorphism.

9. Hypothesis

I hypothesized that there would be genetic factors that affect the development of thyroid cancer in Korean and genetic susceptibility loci in Korea would be different from western countries. Because the epidemiological characteristics of thyroid cancer in Korean are quite different from those in the European descent. Moreover, it is presumed that there is a genetic factor involved in thyroid nodules because of high degree of heritability of thyroid nodule. And thyroid cancer and thyroid nodules may also be genetically related.

10. Aims of study

In Chapter I, I aimed to identify potential susceptibility loci for thyroid cancer, specially PTC and FTC, using GWAS in Korea, and conducted expression quantitative trait loci analysis to validate candidate SNPs using RNA-sequencing data from thyroid cancer and normal tissues. Moreover, I evaluated whether identified signals could have associations with clinical phenotype such as tumour aggressiveness and poor prognosis.

In Chapter II, I conduct GWAS and expression quantitative trait loci analysis to identify susceptibility loci for thyroid nodules. And I also compare them with those for thyroid cancer in a Korean population.

Chapter I. Genome-wide association and expression quantitative trait loci studies for thyroid cancer

Materials and methods

Study participants for the Stage 1 genome scan

DNA samples of thyroid cancer cases for the Stage 1 genome scan were collected in Seoul National University Hospital. These cases comprised of 470 DTC patients (410 PTC and 60 FTC), who underwent thyroidectomy. The baseline characteristics of the study subjects are shown in Table 2. The controls were comprised of 8,279 subjects and recruited from KoGES (Korean Genome and Epidemiology Study), Ansung or Ansan cohort, of which initial investigation began in 2001 with 8,842 participants aged 40–69 (43). The controls were not evaluated on thyroid disease. All participants in this study were of Korean ancestry (Figure. 1).

Study participants for the Stage 2 follow-up

For validation of the candidate associations, we used independent case-control groups. The cases comprised 615 subjects; 524 of the samples (515 PTC and 9 FTC) were from the National Cancer Center, and 91 of the samples (72 PTC and 19 FTC) were from Seoul National University Hospital. Six hundred and five controls were taken from the National Cancer Center and Seoul National University Hospital, respectively (Figure. 1). The DNA samples from the cases were collected at the time of thyroidectomy, and those from the controls were collected when they underwent a health check-up. All of the controls showed a normal thyroid in the ultrasonography examination or had pathologically proven benign nodules. For the eQTL study, 78 thyroid cancer cases (60

PTC and 18 FTC) that had RNA-sequencing results of their tumour or normal tissues were enrolled from Seoul National University Hospital. The baseline characteristics of the cases with thyroid cancer are shown in Table 2. All of the subjects in the replication study were residents of Korea and of Korean ancestry.

Table 2. Descriptive characteristics of the participants in GWAS of DTC

Characteristics	Total (joint)	Stage 1 (discovery)	Stage 2 (replication)
Cases			
Number	1085	470	615
Age, years ± SD	46.5 ± 12.1	43.9 ± 12.8	48.4 ± 11.1
Male %	14.4 %	13.0 %	15.4 %
Pathology			
PTC: FTC, n (%)	997:88 (91.9:8.1)	410:60 (87.2:12.8)	587:28 (95.4:4.6)
$BRAF^{V600E}$ in PTC (%)	186/215 (86.5)	186/215 (86.5)	-
LN metastasis in PTC (%)	419/827 (50.7)	186/337 (55.2)	233/490 (47.6)
ETE in PTC (%)	519/886 (58.6)	221/377(58.6)	298/509 (58.5)
Distant metastasis in PTC (%)	5/769 (0.7)	2/338 (0.6)	3/431 (0.7)
Recurrence in PTC (%)	94/851 (11.0)	54/389 (13.9)	40/462 (8.7)
Controls			
Number	8884	8279	605
Age, years ± SD	52.6 ± 8.8	52.1 ± 8.9	58.9 ± 4.3
Male %	47.5 %	45.2 %	79.2 %

ETE, extrathyroidal extension; LN, lymph node; SD, standard deviation.

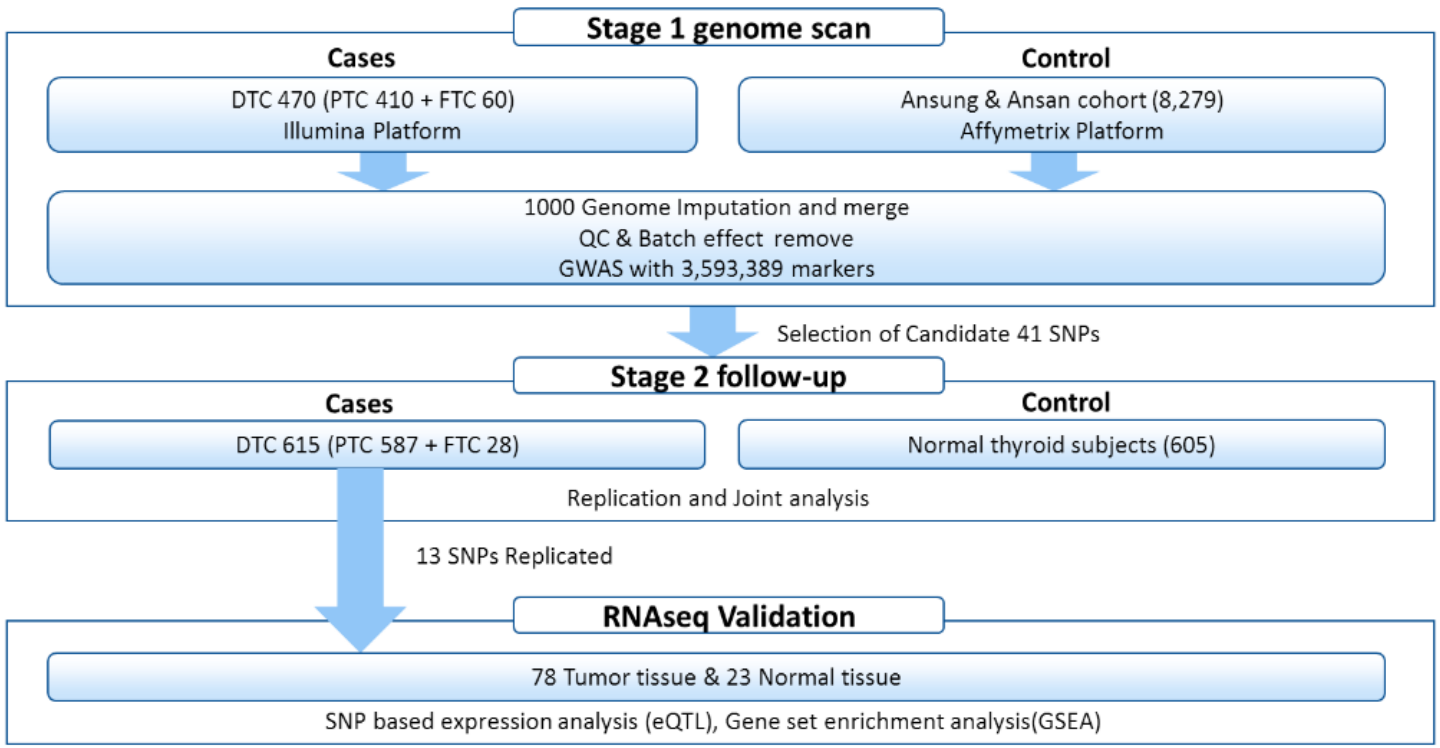

Figure 1. Overview of study flow. The number of individuals with DTC (PTC + FTC) or unaffected individual, imputed or replicated genotypes are shown for each study flow.

Discovery SNP genotyping and imputation

The DNA was extracted from leukocytes of peripheral blood samples obtained from study individuals. For the stage 1 genome scan, the thyroid cancer case samples were genotyped using the Illumina HumanCore-24 BeadChip kit (Illumina, San Diego, USA) and control samples were genotyped using the Affymetrix Genome-Wide Human SNP Array 5.0 (Affymetrix Inc., Santa Clara, USA) (44). To minimize the possible genotyping errors, The SNPs were excluded by the criteria defined by Hardy-Weinberg equilibrium ($P < 1 \times 10^{-6}$), call rate (< 95%), and minor allele frequency (< 1%). After strand alignment with PLINK v1.9 and phasing with SHAPEIT2, imputation was carried out using IMPUTE2 software in both cases and controls. The 1000 genome ASN Phase I reference panel (integrated variant set release in NCBI build 37, hg19) was used as a reference. For imputation quality control, only variants with Info Score < 0.9 were excluded. After merging datasets from cases and controls, we excluded SNPs with a missing genotype rate ≥ 5% (1,590,137 SNPs excluded), and SNPs whose genotype frequencies were out of range from Hardy-Weinberg equilibrium at $P < 1 \times 10^{-6}$ (290 SNPs excluded). Because we used two different platforms for genotyping (Illumina for cases and Affymetrix for controls), there would be spurious associations due to batch effect (45). To detect false associations, we used SNPs data from another healthy cohort comprised of

2,000 subjects, which was genotyped using Illumina HumanCore-24 BeadChip. We tested for batch effects by analyzing association between SNPs from controls and healthy cohorts. We selected for *P*-values using a conservative threshold of 1×10^{-3}, and a total of 31,279 SNPs were excluded. Finally, 3,593,389 markers were used for selecting candidate SNPs.

Replication SNP selection and genotyping

Total forty-one independent variants were selected for the replication test. Among the candidate SNPs with $P < 2.0 \times 10^{-5}$ from the genome-wide scan, we selected representing 27 SNPs in 25 candidate regions base on the clustering pattern and LD. In addition, we included 14 SNPs, which located in previously reported risk loci of thyroid cancer or thyroid disease (Table 2).

For the stage 2 follow-up genotyping, the selected SNPs were genotyped using the Fluidigm SNP Type Assay platform (Fluidigm, San Francisco, USA). To maintain the genotyping quality, a genotyping call rate of > 95% and a Hardy-Weinberg equilibrium with a $P > 0.001$ were considered.

Table 3. Forty-one candidate SNPs for stage2 follow-up study.

Chr.	SNP	Position	Genes	Risk alelle	Allele frequency in cases	Allele frequency in controls	OR	*P*-value
1	rs57075645	99862185	87kb 3' of *LPPR4*	A	0.101	0.060	1.76	3.81E-07
1	rs4915076	108359505	Intronic *VAV3*	C	0.233	0.301	0.71	9.37E-06
1	rs4649295	233416538	Intronic *PCNXL2*	T	0.124	0.180	0.64	1.04E-05
2	rs2121260	38268763	intronic *FAM82A1*	G	0.169	0.125	1.43	8.34E-05
2	rs1979142	39032703	intronic *DHX57*	G	0.187	0.132	1.51	1.76E-06
2	rs1549738	218118722	30kb 3' of *DIRC3*	G	0.388	0.448	0.78	2.96E-04
2	rs12990503	218294217	intronic *DIRC3*	G	0.315	0.375	0.76	1.82E-04
3	rs9858271	59545330	190kb 3' of *FHIT*	G	0.503	0.426	1.37	3.57E-06
4	rs1874564	77858105	13kb 5' of *SEPT11*	A	0.233	0.305	0.69	3.43E-06
4	rs6841841	182790626	270kb 3' of *MGC45800*	G	0.166	0.232	0.66	3.68E-06
5	rs10447240	119328679	357kb 3' of *FAM170A*	A	0.082	0.047	1.36	2.26E-05
6	rs16889600	78723162	550kb 5' of *HTR1B*	T	0.556	0.484	1.83	9.34E-07
6	rs9361385	78926958	650kb 5' of *IRAK1BP1*	C	0.043	0.022	1.34	1.57E-05
6	rs11754852	92261591	30kb 3' of *MIR4643*	C	0.013	0.049	1.99	3.83E-05
7	rs2952745	52196144	812kb 5' of *COBL*	T	0.372	0.302	0.25	4.74E-07
7	rs2715152	82457666	intronic *PCLO*	G	0.069	0.116	1.37	6.24E-06
8	rs36041430	24199218	missense *ADAM28*	A	0.324	0.253	0.57	1.72E-05
8	rs12542743	32318355	intronic *NRG1*	C	0.300	0.225	1.42	1.12E-06
8	rs6996585	32400803	intronic *NRG1*	G	0.261	0.190	1.48	1.20E-07
8	rs2439302	32432369	intronic *NRG1*	G	0.269	0.213	1.36	8.38E-05
8	rs11778356	55387405	14kb 3' of *SOX17*	A	0.284	0.228	1.51	1.20E-07
9	rs4628781	18795997	intronic *ADAMTSL1*	C	0.070	0.038	1.35	6.81E-05
9	rs10867527	83023162	682kb 3' of *TLE4*	G	0.115	0.067	1.93	6.62E-07
9	rs1588635	100537802	78kb 5' of *FOXE1*	A	0.168	0.123	1.81	1.74E-08
9	rs7028661	100538470	77kb 5' of *FOXE1*	A	0.115	0.067	1.8	2.52E-08
9	rs965513	100556109	59kb 5' of *FOXE1*	A	0.109	0.061	1.91	2.35E-09
9	rs1867277	100615914	5'-UTR *FOXE1*	A	0.116	0.075	1.44	4.38E-05
9	rs10122541	100628268	9kb 3' of *FOXE1*	G	0.121	0.080	1.57	1.43E-05
9	rs7037324	100658318	9kb 3' of *C9orf156*	A	0.121	0.080	1.56	1.70E-05
9	rs72753537	100660746	6kb 3' of *C9orf156*	C	0.030	0.012	1.63	3.56E-06
11	rs67790686	103885141	Intronic *PDGFD*	C	0.207	0.153	2.50	3.46E-06
12	rs11175834	65992636	132kb 3' of *MSRB3*	T	0.052	0.022	1.45	1.16E-05
12	rs16934253	113737225	3'-UTR *SLC24A6*	A	0.203	0.146	2.46	2.49E-09
12	rs11061290	131518747	intronic *GPR133*	T	0.134	0.100	1.50	1.41E-06
13	rs75150143	52428825	7.3kb 5' of *CCDC70*	C	0.474	0.414	1.40	8.18E-04
14	rs34081947	36559531	208kb 3' of *MBIP*	T	0.511	0.457	1.28	2.40E-04
14	rs944289	36649246	119kb 3' of *MBIP*	T	0.464	0.406	1.24	1.41E-03
14	rs72693081	81453862	Intronic *TSHR*	G	0.429	0.364	1.27	4.37E-04
19	rs7248104	7224431	Intronic *INSR*	A	0.114	0.072	1.31	6.77E-05
22	rs7288885	26408660	Intornic *MYO18B*	G	0.112	0.075	1.65	2.42E-06

The SNP positions are indexed to the National Center for Biotechnology Information (NCBI) build 37. SNPs with *P*-values > 1×10^{-5} were previous reported SNPs or candidate SNPs of previous reported regions.

Chr, chromosome number; OR, odd ratio; SNP, single nucleotide polymorphism.

RNA sequencing and eQTL analysis

Details of the RNA sequencing methods used have been previously reported (46). In brief, 78 tumor tissues and 23 normal tissues from case samples of replication stage were sequenced using a HiSeq 2000 platform (Illumina, San Diego, USA). Then we profiled gene expression according to Ensembl gene set with the count number of reads aligned to each gene using HTSeq-count and normalized them via fragments per kilobase of exon per million fragments mapped (FPKM). To investigate the *cis*-eQTL of chosen SNPs to neighboring genes (±500 kb), RNA sequencing profiled data was assessed according to the additive model of linear regression analysis. In addition, we evaluated effect of associated genotypes on expression in various tissues using public eQTL database (GTEx (47) and Whole blood eQTL (48)). In patients with PTC from discovery stage, several aggressive features including presence of $BRAF^{V600E}$ mutation, lymph node metastasis, extrathyroid extension and recurrence were analyzed according to genotypes of GWAS-identified variants.

Statistical analysis

We conducted the case-control association analysis with the genome-wide SNPs via an additive model using PLINK software, version 1.9 (https://www.cog-genomics.org/plink/1.9) (49). Logistic regression analysis

was used to test the association for the series of GWAS, replication and joint analysis with unadjusted models, as well as with the adjustment for age and sex. To eliminate relatedness between each pair of subjects in the Stage 1 genome scan, kinship identical-by-descent (IBD) coefficient ($Z0 > 0.8$) was considered. Q-Q plots were used to assess the adequacy of the case-control matching. We also calculated the genomic inflation factor (λ) from a GWA analysis to compare the genome-wide distribution of the test statistic with the expected null distribution. The regional plots were created using LocusZoom (http://locuszoom.sph.umich.edu/locuszoom). The gene set enrichment analysis was conducted using GSEA software, version 2.2.1 (Broad Institute, www.broad.mit.edu/gsea/msigdb/index.jsp) with the BioCarta, KEGG and Reactome (1077 gene sets) of Molecular Signatures Database (MSigDB version 5.1, http://www.broadinstitute.org/gsea/msigdb/) (50). The IBM SPSS Statistics for Windows, version 23 (IBM Corp., Armonk, USA) was used for the statistical analyses.

Ethics statement

This study was approved by the Institutional Review Board of Seoul National University Hospital (IRB No. H-1102-012-349 and H-1108-041-372) and National Cancer Center (IRB No. NCC2015-0027), and written informed consents were obtained from all participants. All clinical investigations were conducted according to the principles expressed in the Declaration of Helsinki.

Results

Stage 1 genome scan

After genotype imputation, quality control, and the removal of the batch effect and relatedness, we conducted an association analysis using 3,593,389 markers for DTC, papillary thyroid cancer (PTC) and follicular thyroid cancer (FTC) versus the control (Figure 1, Table 2). A quantile-quantile (Q-Q) plot and genomic inflation factors showed little evidence for statistic inflation (Figure. 2). The genome-wide association results of each of the DTC and PTC cases are shown in the Manhattan plots (Figure 3). We identified two genome-wide significant ($P = 5 \times 10^{-8}$) loci in DTC and two loci in PTC. In the test using the DTC cases, the most significantly associated SNP was observed near the *FOXE1* gene (rs965513, $P = 2.35 \times 10^{-9}$) at 9q22.33. The second significant SNP was located in the 3'-untranslated region (UTR) of the *SLC24A6* gene (rs16934253, $P = 2.49 \times 10^{-9}$) at 12q24.13 (Figure. 1a). In the test using the PTC cases, we found two significantly associated signals in the intronic region of the *NRG1* gene (rs6996585, $P = 5.17 \times 10^{-10}$) at 8p12 and near the *FOXE1* gene (rs965513, $P = 8.20 \times 10^{-9}$), which was the most significant SNP in DTC (Figure 1b). Forty-one candidate SNPs were selected for the replication (Table 2). Among these, only six SNPs of *DIRC3, NKX2-1* or *NRG1* were identical to the previous reports (Table 1).

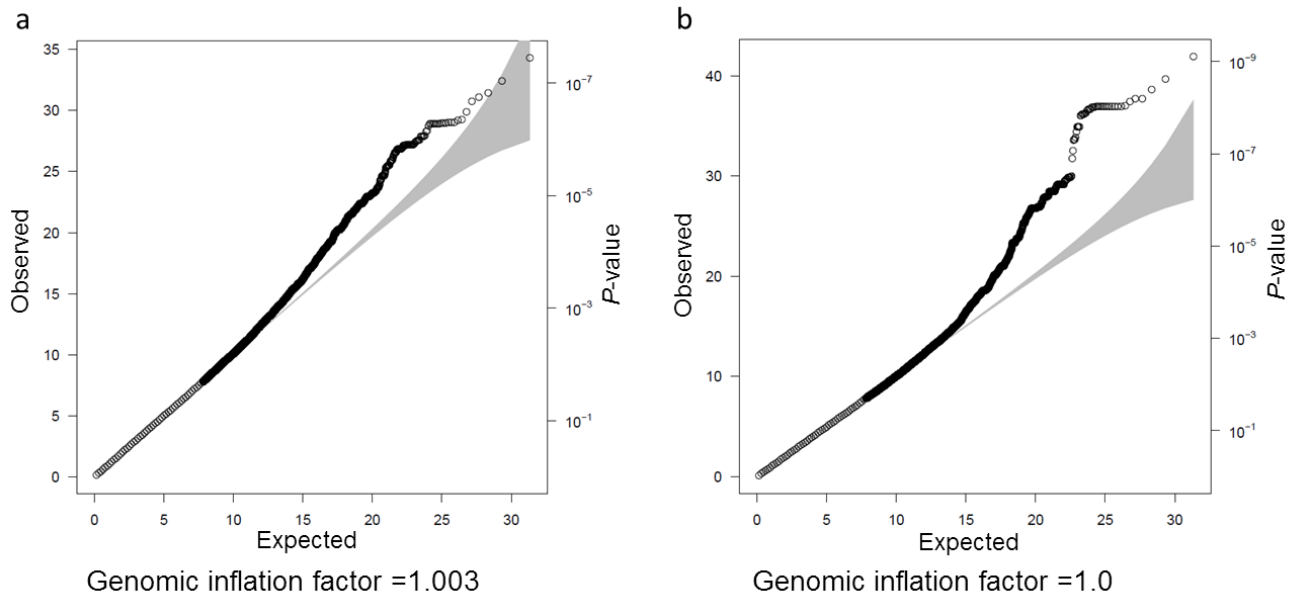

Figure 2. Quantile-quantile plot for stage 1 genome scan.

A quantile-quantile plot for (a) DTC and (b) PTC showing the distribution of the observed *P*-values from the association testing in the stage 1 genome scan against the expected distribution under the null hypothesis. The grey zone indicates the 95% confidence interval.

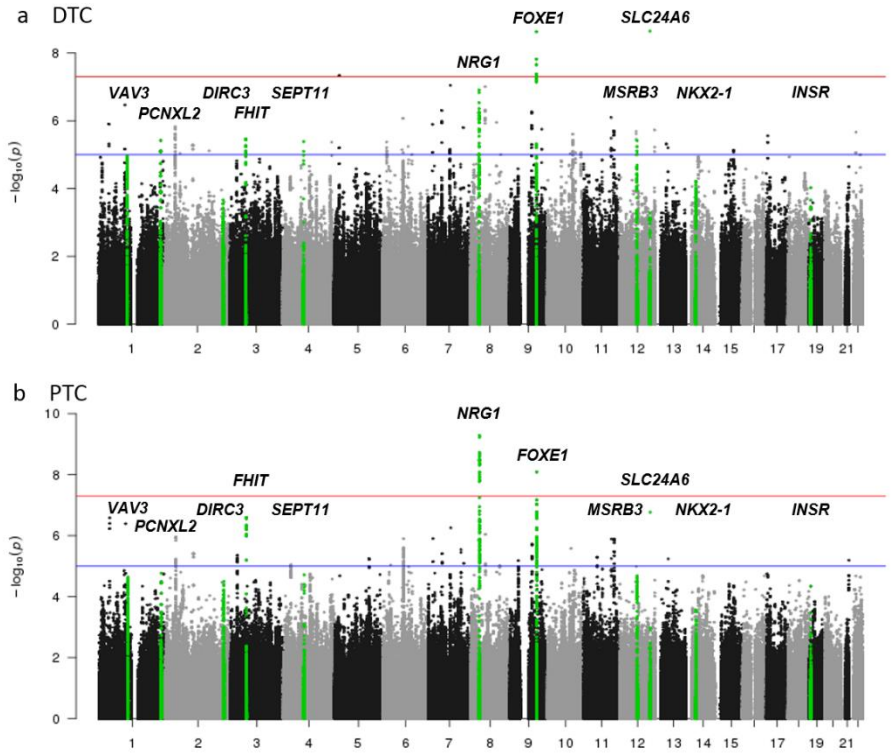

Figure 3. Manhattan plots of the Genome-wide association signal with DTC and PTC for stage 1. The x-axis represents the SNP markers on each chromosome. y-axis shows the $-\log_{10}$ scale. The red horizontal line represents the genome-wide significant threshold $P = 5.0 \times 10^{-8}$ and the blue horizontal line represents the genome-wide suggestive threshold $P = 1.0 \times 10^{-6}$. Eleven candidate loci of DTC, PTC or FTC are shown in green.

Stage 2 follow-up and joint Stages 1 and 2 analyses

Among the 41 candidate SNPs, 13 SNPs in 10 loci were replicated in DTC or PTC, and 2 SNPs in 2 loci were replicated in FTC (green coloured loci in Figure 1). The minor allele frequency (MAF) of all of the 15 SNPs in our control samples were similar with those of the East Asian population in 1000 genomes (Table 4), with the exception of rs1549738. Table 5 describes the replicated SNPs in DTC, PTC and FTC through the GWAS and the replication and joint association analysis. Most of the SNPs showed a similar association between DTC and PTC, with the exception of 2 SNPs; rs4915076 of *VAV3* and rs9858271 of *FHIT* were replicated only in PTC. After the joint analysis, the most significantly associated region was the *NRG1* loci, and the second one was *DIRC3*.

Regarding the candidate SNPs that were associated with FTC, despite the limited number of FTC samples (discovery $N = 60$, replication $N = 28$), we identified two SNPs that were highly associated with FTC but were not positively replicated in the DTC or PTC samples. The SNP (rs16934253) in the 3' UTR of *SLC24A6* at 12q24.13 is the second most significantly associated signal of DTC in the discovery stage, but it was not well replicated in DTC or PTC. However, we identified that rs16934253 showed a high-risk effect (Joint

P = 2.71 × 10^{-5}, OR = 3.32) in the FTC samples. Another candidate SNP (rs1549738) near *DIRC3* showed a similar risk effect in the FTC samples (Joint P = 0.0017, OR = 1.65) but not in the DTC or PTC samples.

Among the SNPs, 2 SNPs in 2 loci were identical (rs2439302 in the *NRG1* locus and rs944289 in the *NKX2-1* locus), and 6 SNPs in 4 loci (rs6996585 and rs12542743 in the *NRG1* locus, rs12990503 and rs1549738 in the *DIRC3* locus, rs34081947 in the *NKX2-1* locus and rs72753537 in the *FOXE1* locus) were located at the same loci that were identified in previous reports (Table 1). The other 7 SNPs in 7 loci (*VAV3, PCNXL2, INSR, MRSB3, FHIT, SEPT11* and *SLC24A6* loci) were newly identified.

Table 4. The comparison of risk allele frequency between population of 1000Genome and this study

Chr	SNP	Gene	Risk / Reference allele	Risk allele frequency in 1000Genome				Risk allele frequency in this study		
				African	American	European	East Asian	Cases	Controls	Allelic OR
1	rs4915076	*VAV3*	T/C	0.96	0.81	0.93	0.71	0.76	0.70	1.33
1	rs4649295	*PCNXL2*	C/T	0.43	0.75	0.64	0.84	0.87	0.82	1.43
2	rs12990503	*DIRC3*	G/C	0.54	0.35	0.28	0.60	0.69	0.63	1.34
2	rs1549738	*DIRC3*	A/G	0.54	0.84	0.87	0.61	0.58	0.55	1.14
3	rs9858271	*FHIT*	G/A	0.07	0.29	0.24	0.47	0.48	0.43	1.26
4	rs1874564	*SEPT11*	G/A	0.33	0.52	0.45	0.66	0.75	0.69	1.31
8	rs6996585	*NRG1*	G/A	0.24	0.45	0.42	0.23	0.29	0.23	1.39
8	rs12542743	*NRG1*	C/T	0.51	0.56	0.56	0.26	0.32	0.25	1.36
8	rs2439302	*NRG1*	G/C	0.47	0.49	0.48	0.19	0.27	0.21	1.37
9	rs72753537	*FOXE1*	C/T	0.04	0.12	0.14	0.08	0.10	0.07	1.41
12	rs11175834	*MSRB3*	T/C	0.40	0.10	0.05	0.14	0.20	0.15	1.37
12	rs16934253	*SLC24A6*	A/G	0.32	0.08	0.11	0.01	0.03	0.02	1.51
14	rs34081947	*NKX2-1*	T/C	0.20	0.43	0.54	0.39	0.47	0.41	1.27
14	rs944289	*NKX2-1*	T/C	0.15	0.44	0.59	0.45	0.51	0.46	1.25
19	rs7248104	*INSR*	A/G	0.28	0.40	0.42	0.32	0.41	0.36	1.22

Chr, chromosome number; OR, odd ratio; SNP, single nucleotide polymorphism.

Table 5. DTC, PTC and FTC associated SNPs in Korean Population

SNP	Chr Position Gene	Risk Allele	Stage	DTC			PTC			FTC		
				Allele Frequency (case/control)	OR	*P*	Allele Frequency (case/control)	OR	*P*	Allele Frequency (case/control)	OR	*P*
rs6996585	8	G	Discovery	0.30/0.23	1.48	1.20E-07	0.32/0.23	1.61	5.17E-10	0.17/0.23	0.70	0.1499
	32400803		Replication	0.28/0.23	1.29	0.0061	0.28/0.23	1.28	0.0094	0.32/0.23	1.59	0.1130
	NRG1		Joint	0.29/0.23	1.39	1.08E-10	0.29/0.23	1.43	9.01E-12	0.22/0.23	0.96	0.8273
rs12542743	8	C	Discovery	0.32/0.25	1.42	1.12E-06	0.34/0.25	1.53	1.63E-08	0.21/0.25	0.78	0.2674
	32318355		Replication	0.31/0.27	1.22	0.0267	0.31/0.27	1.20	0.0427	0.38/0.27	1.63	0.0821
	NRG1		Joint	0.32/0.25	1.36	4.61E-10	0.32/0.25	1.39	1.01E-10	0.26/0.25	1.04	0.8137
rs2439302	8	G	Discovery	0.27/0.21	1.36	8.38E-05	0.29/0.21	1.48	1.50E-06	0.15/0.21	0.66	0.1129
	32432369		Replication	0.27/0.21	1.37	8.55E-04	0.27/0.21	1.36	0.0013	0.30/0.21	1.59	0.1143
	NRG1		Joint	0.27/0.21	1.37	1.42E-09	0.28/0.21	1.41	1.26E-10	0.20/0.21	0.94	0.7289
rs12990503	2	G	Discovery	0.68/0.62	1.32	1.82E-04	0.70/0.62	1.38	3.17E-05	0.61/0.62	0.93	0.7107
	218294217		Replication	0.70/0.65	1.21	0.0268	0.70/0.65	1.24	0.0164	0.61/0.65	0.83	0.5119
	DIRC3		Joint	0.69/0.63	1.34	3.55E-09	0.70/0.63	1.38	2.58E-10	0.61/0.63	0.93	0.6324
rs11175834	12	T	Discovery	0.21/0.15	1.45	1.16E-05	0.21/0.15	1.44	4.52E-05	0.21/0.15	1.48	0.0879
	65992636		Replication	0.19/0.14	1.41	0.0018	0.19/0.14	1.38	0.0035	0.25/0.14	1.99	0.0289
	MSRB3		Joint	0.20/0.15	1.37	4.26E-08	0.20/0.15	1.36	4.86E-07	0.22/0.15	1.60	0.0100
rs4915076	1	T	Discovery	0.77/0.70	1.42	9.37E-06	0.77/0.70	1.41	4.51E-05	0.77/0.70	1.48	0.0707
	108359505		Replication	0.75/0.71	1.20	0.0507	0.75/0.71	1.23	0.0240	0.63/0.71	0.68	0.1637
	VAV3		Joint	0.76/0.70	1.33	8.47E-08	0.76/0.70	1.34	7.09E-08	0.73/0.70	1.14	0.4311
rs4649295	1	T	Discovery	0.88/0.82	1.56	1.04E-05	0.88/0.82	1.56	3.48E-05	0.87/0.82	1.54	0.1155
	233416538		Replication	0.86/0.82	1.33	0.0106	0.86/0.82	1.36	0.0068	0.80/0.82	0.89	0.7326
	PCNXL2		Joint	0.87/0.82	1.43	6.00E-08	0.87/0.82	1.45	8.53E-08	0.85/0.82	1.27	0.2634
rs34081947	14	T	Discovery	0.47/0.41	1.28	2.40E-04	0.47/0.41	1.24	0.003163	0.53/0.41	1.62	0.0079
	36559531		Replication	0.47/0.39	1.38	8.07E-05	0.47/0.39	1.37	1.31E-04	0.50/0.39	1.56	0.0999
	NKX2-1		Joint	0.47/0.41	1.27	1.19E-07	0.47/0.41	1.25	2.47E-06	0.52/0.41	1.56	0.0030

SNP	Chr Position Gene	Risk Allele	Stage	DTC			PTC			FTC		
				Allele Frequency (case/control)	OR	*P*	Allele Frequency (case/control)	OR	*P*	Allele Frequency (case/control)	OR	*P*
rs1874564	4	G	Discovery	0.77/0.69	1.44	3.43E-06	0.77/0.69	1.43	1.93E-05	0.77/0.69	1.51	0.0576
	77858105		Replication	0.74/0.69	1.24	0.0169	0.74/0.69	1.25	0.0135	0.69/0.69	1.00	0.9930
	SEPT11		Joint	0.75/0.69	1.31	2.04E-07	0.75/0.69	1.31	5.87E-07	0.75/0.69	1.32	0.1171
rs9858271	3	G	Discovery	0.50/0.43	1.37	3.57E-06	0.52/0.43	1.45	2.78E-07	0.41/0.43	0.92	0.6570
	59545330		Replication	0.47/0.43	1.15	0.0952	0.47/0.43	1.18	0.0468	0.32/0.43	0.62	0.0989
	FHIT		Joint	0.48/0.43	1.26	6.82E-07	0.49/0.43	1.30	2.76E-08	0.38/0.43	0.82	0.2029
rs944289	14	T	Discovery	0.51/0.46	1.24	0.0014	0.51/0.46	1.22	0.0062	0.54/0.46	1.40	0.0646
	36649246		Replication	0.51/0.43	1.38	7.53E-05	0.51/0.43	1.36	1.93E-04	0.59/0.43	1.90	0.0186
	NKX2-1		Joint	0.51/0.46	1.25	1.39E-06	0.51/0.46	1.23	1.72E-05	0.56/0.46	1.50	0.0072
rs72753537	9	C	Discovery	0.12/0.07	1.63	3.56E-06	0.12/0.07	1.76	1.70E-07	0.06/0.07	0.77	0.4958
	100660746		Replication	0.09/0.07	1.38	0.0352	0.09/0.07	1.43	0.0209	0.04/0.07	0.52	0.3560
	FOXE1		Joint	0.10/0.07	1.41	7.67E-06	0.11/0.07	1.48	5.37E-07	0.05/0.07	0.67	0.2448
rs7248104	19	A	Discovery	0.43/0.36	1.31	6.77E-05	0.43/0.36	1.35	4.57E-05	0.39/0.36	1.11	0.5877
	7224431		Replication	0.40/0.35	1.20	0.0313	0.40/0.35	1.20	0.0293	0.38/0.35	1.10	0.7449
	INSR		Joint	0.41/0.36	1.22	2.00E-05	0.41/0.36	1.23	1.64E-05	0.38/0.36	1.09	0.5731
FTC associated SNPs				**DTC**			**PTC**			**FTC**		
rs16934253	12	A	Discovery	0.05/0.02	2.46	2.49E-09	0.05/0.02	2.36	1.71E-07	0.07/0.02	3.20	8.95E-04
	113737225		Replication	0.02/0.02	0.98	0.9573	0.02/0.02	0.83	0.5752	0.07/0.02	4.35	0.0045
	SLC24A6		Joint	0.03/0.02	1.51	0.0016	0.03/0.02	1.36	0.0216	0.07/0.02	3.32	2.71E-05
rs1549738	2	A	Discovery	0.61/0.55	1.28	2.96E-04	0.61/0.55	1.25	0.0026	0.66/0.55	1.56	0.0193
	218118722		Replication	0.56/0.56	1.04	0.6595	0.56/0.56	1.01	0.9003	0.70/0.56	1.84	0.0376
	DIRC3		Joint	0.58/0.55	1.14	0.0036	0.58/0.55	1.11	0.0307	0.67/0.55	1.65	0.0017

The SNP positions are indexed to the National Center for Biotechnology Information (NCBI) build 37.
Chr, chromosome number; OR, odd ratio; SNP, single nucleotide polymorphism.

Validation of the candidate SNPs with cis-eQTL and GSEA analyses.

We conducted a *cis*-eQTL analysis with the 15 positively replicated SNPs and their nearby genes using 78 tumour and 23 normal thyroid tissues from samples of the replication stage (Table 6). We found significant *cis*-eQTL of genes near *NRG1, NKX2-1, DIRC3, PCNXL2* and *VAV3* in the normal and tumour thyroid tissue. Additionally, we evaluated the *cis*-eQTL of the normal thyroid tissues (Table 7) and other various tissues (Table 8) in the public eQTL database and imputed expression (Table 6) of 470 DTC case samples using PrediXcan, found similar results with our RNAseq eQTL data. To reveal the transcriptional change and biological function during the cancer prognosis, we conducted a gene set enrichment analysis (GSEA) according to the candidate SNPs in the tumour and normal thyroid tissue. We found several cellular growths or cancer-related pathways that were associated with SNPs of *NRG1, VAV3, DIRC3, SEPT11* and *INSR* (Tables 9). Specifically, rs6996585 of *NRG1* was observed in association with a number of those pathways in the normal thyroid tissue (Tables 10).

Table 6. *Cis*-eQTL result of candidate SNPs and nearby genes.

Chr	SNP	Position	Representative Gene	RNA-sequencing data in this study			Predicted expression of 470 DTC cases using normal thyroid reference	
				Cis-eQTL Gene	*P*-value of tumour tissue	*P*-value of normal tissue	*Cis*-eQTL Gene	*P*-value
1	rs4915076	108359505	*VAV3*	*VAV3*	**0.0174**	0.0995	*VAV3*	**< 1.00E-300**
1	rs4649295	233416538	*PCNXL2*	*PCNXL2*	**0.0030**	0.8594	*PCNXL2*	**3.81E-05**
				NTPCR	0.9006	**0.0472**	*NTPCR*	**0.0262**
2	rs1549738	218118722	*DIRC3*	*TNS1*	**0.0023**	0.1170	*TNS1*	NA
2	rs12990503	218294217	*DIRC3*	-			-	
3	rs9858271	59545330	*FHIT*	-			-	
4	rs1874564	77858105	*SEPT11*	-			-	
8	rs6996585	32400803	*NRG1*	*NRG1*	**0.0053**	0.0526	*NRG1*	**2.99E-244**
8	rs12542743	32318355	*NRG1*	*NRG1*	**0.0073**	0.1021	*NRG1*	**3.04E-97**
8	rs2439302	32432369	*NRG1*	*NRG1*	**0.0025**	**0.0125**	*NRG1*	**< 1.00E-300**
9	rs72753537	100660746	*FOXE1*	*C9orf156*	0.6914	0.3035	*C9orf156*	**1.09E-43**
				CORO2A	0.0551	0.4374	*CORO2A*	**1.03E-19**
				XPA	0.4269	0.9955	*XPA*	**8.43E-58**
				TSTD2	0.3061	0.2419	*TSTD2*	**0.0013**
12	rs11175834	65992636	*MSRB3*	-			-	
12	rs16934253	113737225	*SLC24A6*	-			-	
14	rs34081947	36559531	*NKX2-1*	*NKX2-1*	**0.0323**	0.5458	*NKX2-1*	NA
				SFTA3	0.0883	0.4173	*SFTA3*	**4.05E-15**
14	rs944289	36649246	*NKX2-1*	*NKX2-1*	**0.0069**	**0.0302**	*NKX2-1*	NA
				SFTA3	**0.0107**	**0.0476**	*SFTA3*	**3.90E-13**
				RALGAPA1	0.2493	0.2766	*RALGAPA1*	**0.0172**
19	rs7248104	7224431	*INSR*	*INSR*	0.7187	0.8680	*INSR*	**4.91E-41**

The SNP positions are indexed to the National Center for Biotechnology Information (NCBI) build 37. The *cis*-eQTL gene is defined as the genes nearest to the candidate SNP within ±500 kb. The *cis*-eQTL of candidate SNPs are from the association result of 78 tumour thyroid tissues and 23 normal thyroid tissues. Bold indicates significance of $P < 0.05$.

Chr, chromosome number; NA, not available; SNP, single nucleotide polymorphism.

Table 7. Association between candidate SNPs and *cis*-eQTL result of thyroid tissues in GTEx public data

Chr	SNP	Position	Representative Gene	RNA-sequencing data in this study			Public data in normal thyroid (source: GTEx2015 v6)	
				Cis-eQTL Gene	*P*-value of tumor tissue	*P*-value of normal tissue	*Cis*-eQTL Gene	*P*-value
1	rs4915076	108359505	*VAV3*	*VAV3*	**0.0174**	0.0995	*VAV3*	**3.33E-27**
							VAV3-AS1	**2.02E-06**
1	rs4649295	233416538	*PCNXL2*	*PCNXL2*	**0.0030**	0.8594	*PCNXL2*	> 0.05
				NTPCR	0.9006	**0.0472**	*NTPCR*	> 0.05
2	rs1549738	218118722	*DIRC3*	*TNS1*	**0.0023**	0.1170	*TNS1*	> 0.05
2	rs12990503	218294217	*DIRC3*	-				
3	rs9858271	59545330	*FHIT*	-				
4	rs1874564	77858105	*SEPT11*	-				
8	rs6996585	32400803	*NRG1*	*NRG1*	**0.0053**	0.0526	*NRG1*	**5.79E-21**
				-			*RP11-1002K11.1*	**6.46E-19**
8	rs12542743	32318355	*NRG1*	*NRG1*	**0.0073**	0.1021	*NRG1*	**2.50E-07**
				-			*RP11-1002K11.1*	**1.00E-06**
8	rs2439302	32432369	*NRG1*	*NRG1*	**0.0025**	**0.0125**	*NRG1*	**6.47E-25**
				-			*RP11-1002K11.1*	**1.76E-23**
9	rs72753537	100660746	*FOXE1*	*C9orf156*	0.6914	0.3035	*C9orf156*	**1.33E-05**
12	rs11175834	65992636	*MSRB3*	-				
12	rs16934253	113737225	*SLC24A6*	-				
14	rs34081947	36559531	*NKX2-1*	*NKX2-1*	**0.0323**	0.5458	*NKX2-1*	> 0.05
				-			*RP11-116N8.4*	**2.90E-12**
				-			*PTCSC3*	**1.50E-05**
14	rs944289	36649246	*NKX2-1*	*NKX2-1*	**0.0069**	**0.0302**	*NKX2-1*	> 0.05
				-			*RP11-116N8.4*	**1.28E-09**
				SFTA3	**0.0107**	**0.0476**	*SFTA3*	> 0.05
19	rs7248104	7224431	*INSR*	-				

The SNP positions are indexed to the National Center for Biotechnology Information (NCBI) build 37. The *cis*-eQTL gene is defined as the genes nearest to the candidate SNP within ±500 kb. The *cis*-eQTL of candidate SNPs are from the association result of 78 tumor thyroid tissues and 23 normal thyroid tissues. The public *cis*-eQTL result of candidate SNPs are from the GTEx (http://www.gtexportal.org). Bold indicates significance of $P < 0.05$.
Chr, chromosome number; SNP, single nucleotide polymorphism.

Table 8. Association between candidate SNPs and *Cis*-eQTL result of other various tissues except thyroid in public data

Chr	SNP	Position	Representative Gene	Source	eQTL-Gene	Tissue	*P*-value
1	rs4915076	108359505	*VAV3*	Westra2013	*VAV3*	Whole blood	1.56E-08
				GTEx2015_v6	*VAV3*	Lung	3.23E-08
				GTEx2015_v6	*VAV3*	Whole blood	1.91E-06
1	rs4649295	233416538	*PCNXL2*	-	-	-	-
2	rs12990503	218294217	*DIRC3*	GTEx2015_v6	*DIRC3*	Skin	7.19E-06
2	rs1549738	218118722	*DIRC3*	-	-	-	-
3	rs9858271	59545330	*FHIT*	-	-	-	-
4	rs1874564	77858105	*SEPT11*	Westra2013	*CCNI*	Whole blood	0.0013
				Westra2013	*SEPT11*	Whole blood	0.0018
8	rs12542743	32318355	*NRG1*	-			
8	rs6996585	32400803	*NRG1*	Westra2013	*NRG1*	Whole blood	3.95E-190
				GTEx2015_v6	*NRG1*	Whole blood	2.49E-11
				GTEx2015_v6	*RP11-1002K11.1*	Whole blood	7.74E-08
8	rs2439302	32432369	*NRG1*	Westra2013	*NRG1*	Whole blood	9.81E-198
				GTEx2015_v6	*NRG1*	Whole blood	1.68E-13
				GTEx2015_v6	*RP11-1002K11.1*	Whole blood	1.68E-09
9	rs72753537	100660746	*FOXE1*	GTEx2015_v6	*C9orf156*	Adipose	1.15E-05
				GTEx2015_v6	*C9orf156*	Skeletal muscle	3.79E-06
				GTEx2015_v6	*C9orf156*	Nerve, Tibia	5.04E-06
				GTEx2015_v6	*C9orf156*	Testis	1.67E-05
12	rs11175834	65992636	*MSRB3*	-	-	-	-
12	rs16934253	113737225	*SLC24A6*	Westra2013	*AC010178.40-2*	Whole blood	3.50E-04
				Westra2013	*C12orf52*	Whole blood	0.0029
				Westra2013	*SLC24A6*	Whole blood	3.67E-26
14	rs34081947	36559531	*NKX2-1*	GTEx2015_v6	*RP11-116N8.4*	Adipose	2.90E-07
14	rs944289	36649246	*NKX2-1*	GTEx2015_v6	*RP11-116N8.4*	Adipose	5.79E-06
19	rs7248104	7224431	*INSR*	GTEx2015_v6	*INSR*	Nerve, Tibia	2.70E-06

The SNP positions are indexed to the National Center for Biotechnology Information (NCBI) build 37. The *cis*-eQTL result of candidate SNPs are from the GTEx (http://www.gtexportal.org) and Whole blood eQTL (Westra 2013). Chr, chromosome number; SNP, single nucleotide polymorphism.

Table 9. Gene set enrichment analysis result in total candidate SNPs

Chr	SNP	Gene	Normal thyroid tissue		Tumor thyroid tissue		Significant gene set list
			N of gene set (FDR q<0.05)	Lowest FDR q	*N* of gene set (FDR q<0.05)	Lowest FDR q	
1	rs4649295	*PCNXL2*	0	0.662	0	0.160	-
1	rs4915076	*VAV3*	0	0.108	**2**	**0.026**	(KEGG) Steroid Hormone Biosynthesis, (Reactome) Steroid Hormones
2	rs12990503	*DIRC3*	0	0.685	**1**	**0.038**	(Reactome) TGF beta receptor signaling activates SMADS
2	rs1549738	*DIRC3*	0	0.267	0	0.200	-
3	rs9858271	*FHIT*	0	0.192	0	0.569	-
4	rs1874564	*SEPT11*	0	0.499	**1**	**0.042**	(Biocarta) ATM Pathway
8	rs12542743	*NRG1*	0	0.234	0	0.860	-
8	rs6996585	*NRG1*	**31**	**< 0.001**	0	0.936	(Biocarta) AT1R Pathway, CXCR4 Pathway, EIF4 Pathway, FCER1 Pathway, FMLP Pathway, GH Pathway, GLEEVEC Pathway, GPCR Pathway, GSK3 Pathway, HCMV Pathway, IGF1 Pathway, IL6 Pathway, Insulin Pathway, MEF2D Pathway, MET Pathway, NFAT Pathway, NFkB Pathway, NGF Pathway, PDGF Pathway, PYK2 Pathway, Stress Pathway, TCR Pathway, VEGF Pathway (KEGG) Axon guidance, Colorectal Cancer (Reactome) Downstream signal trransduction, NGF signalling via TRKA from the plasma membrane, Regulation of KIT signaling, Signaling by FGFR, Signaling by NGF, Transcriptional regulaiton of white adipocyte differntiation
8	rs2439302	*NRG1*	0	0.074	0	0.761	-
9	rs72753537	*FOXE1*	0	0.152	0	0.400	-
12	rs11175834	*MSRB3*	0	0.686	0	0.810	-
12	rs16934253	*SLC24A6*	NA	NA	0	0.997	-
14	rs34081947	*NKX2-1*	0	0.110	0	0.529	-
14	rs944289	*NKX2-1*	0	0.803	0	0.059	-
19	rs7248104	*INSR*	**1**	**0.028**	0	0.650	(Biocarta) ERK Pathway

1077 gene sets (BioCarta, KEGG and Reactome) of Molecular Signatures Database (MSigDB version 5.1) were used. Bold indicates significance of FDR $q < 0.05$.
Chr, chromosome number; FDR q, false discovery rate q-value; OR, odd ratio; SNP, single nucleotide polymorphism.

Table 10. The significantly enriched gene sets (FDR $q < 0.05$) according to the rs6996585 genotype.

Gene set	Description	NES	Nor *P*	FDR *q*
(Biocarta) AT1R Pathway	Angiotensin II mediated activation of JNK Pathway via Pyk2 dependent signaling	2.316	< 0.001	0.001
(Biocarta) PYK2 Pathway	Links between Pyk2 and Map Kinases	2.273	< 0.001	0.004
(Biocarta) IGF1 Pathway	IGF-1 Signaling Pathway	2.108	< 0.001	0.018
(Biocarta) Insulin Pathway	Insulin Signaling Pathway	2.044	< 0.001	0.02
(Biocarta) MET Pathway	Signaling of Hepatocyte Growth Factor Receptor	2.054	< 0.001	0.02
(Reactome) Transcriptional regulation of white adipocyte differentiation	Genes involved in Transcriptional Regulation of White Adipocyte Differentiation	2.124	< 0.001	0.021
(Biocarta) NGF Pathway	Nerve growth factor pathway	2.06	< 0.001	0.022
(Biocarta) GLEEVEC Pathway	Inhibition of Cellular Proliferation by Gleevec	2.084	< 0.001	0.022
(Biocarta) GSK3 Pathway	Inactivation of Gsk3 by AKT causes accumulation of b-catenin in Alveolar Macrophages	2.022	0.002	0.024
(Biocarta) PDGF Pathway	PDGF Signaling Pathway	2.026	< 0.001	0.024
(Biocarta) FCER1 Pathway	Fc Epsilon Receptor I Signaling in Mast Cells	2.002	< 0.001	0.029
(Reactome) NGF signaling via TRKA	Genes involved in NGF signaling via TRKA from the plasma membrane	1.968	< 0.001	0.041
(Biocarta) IL6 Pathway	IL 6 signaling pathway	1.93	< 0.001	0.043
(Biocarta) GPCR Pathway	Signaling Pathway from G-Protein Families	1.931	< 0.001	0.044
(Reactome) Downstream signal transduction	Genes involved in Downstream signal transduction	1.936	< 0.001	0.044
(Biocarta) CXCR4 Pathway	CXCR4 Signaling Pathway	1.95	< 0.001	0.045
(Biocarta) MEF2D Pathway	Role of MEF2D in T-cell Apoptosis	1.938	< 0.001	0.045
(Biocarta) NFkB Pathway	NF-kB Signaling Pathway	1.941	< 0.001	0.046
(KEGG) Colorectal Cancer	Colorectal cancer	1.826	0.002	0.048
(Biocarta) EIF4 Pathway	Regulation of eIF4e and p70 S6 Kinase	1.837	< 0.001	0.048
(Biocarta) HCMV Pathway	Human Cytomegalovirus and Map Kinase Pathways	1.813	0.002	0.048
(KEGG) Axon guidance	Axon guidance	1.839	0.008	0.048
(Biocarta) Stress Pathway	TNF/Stress Related Signaling	1.822	0.008	0.048
(Biocarta) GH Pathway	Growth Hormone Signaling Pathway	1.951	0.008	0.048
(Biocarta) NFAT Pathway	NFAT and Hypertrophy of the heart (Transcription in the broken heart)	1.912	< 0.001	0.049
(Biocarta) TCR Pathway	T Cell Receptor Signaling Pathway	1.837	< 0.001	0.049
(Reactome) Signaling by FGFR	FGFR Signaling pathway	1.816	0.004	0.049
(Reactome) Signaling by NGF	Genes involved in Signaling by NGF	1.839	0.002	0.049
(Biocarta) FMLP Pathway	fMLP induced chemokine gene expression in HMC-1 cells	1.828	0.004	0.049
(Biocarta) VEGF Pathway	VEGF, Hypoxia, and Angiogenesis	1.833	< 0.001	0.049
(Reactome) Regulation of KIT signaling	Genes involved in Regulation of KIT signaling	1.817	0.004	0.049

1077 gene sets (BioCarta, KEGG and Reactome) of Molecular Signatures Database (MSigDB version 5.1) were used.
FDR *q*, false discovery rate *q*-value; NES, normalized enrichment score; Nor *P*, nominal *P*-value.

Association between candidate SNPs and clinical phenotypes.

We further investigated whether there was an association between candidate SNPs and clinical phenotypes, such as the $BRAF^{V600E}$ mutation, lymph node metastasis, or extrathyroidal extension (Table 11). Interestingly, three SNPs of *NRG1* were associated with lymph node metastasis in the $BRAF^{V600E}$ positive samples. The SNPs of *SEPT11* and *INSR* were associated with extrathyroidal extension.

Table 11. Association between candidate SNPs and clinical phenotypes.

Chr	SNP	Gene	Clinical phenotype	Genotypes			Total	*P*-value	*BRAF* positive	*BRAF* negative
1	rs4915076	*VAV3*		CC	C**T**	**TT**				
			BRAF^{V600E}	10/11(90.9%)	73/83(88.0%)	103/121(85.1%)	186/215(86.5%)	0.477		
			LN metastasis	11/17(64.7%)	71/124(57.3%)	104/196(53.1%)	186/337(55.2%)	0.289	0.855	0.653
			ETE	10/20(50.0%)	78/134(58.2%)	133/223(59.6%)	221/377(58.6%)	0.483	0.634	0.573
1	rs4649295	*PCNXL2*		**CC**	**C**T	TT				
			BRAF^{V600E}	141/162(87.0%)	40/46(87.0%)	4/6(66.7%)	185/214(86.4%)	0.377		
			LN metastasis	145/256(56.6%)	34/72(47.2%)	3/4(75.0%)	182/332(54.8%)	0.356	0.930	0.765
			ETE	169/285(59.3%)	46/80(57.5%)	3/7(42.9%)	218/372(58.6%)	0.489	0.302	0.107
2	rs12990503	*DIRC3*		**CC**	**C**G	GG				
			BRAF^{V600E}	94/105(89.5%)	71/87(81.6%)	20/22(90.9%)	185/214(86.4%)	0.491		
			LN metastasis	94/162(58.0%)	69/137(50.4%)	23/37(62.2%)	186/336(55.4%)	0.763	0.928	0.041
			ETE	108/183(59.0%)	85/153(55.6%)	27/40(67.5%)	220/376(58.5%)	0.669	0.584	0.290
2	rs1549738	*DIRC3*		**AA**	**A**G	GG				
			BRAF^{V600E}	65/74(87.8%)	98/113(86.7%)	21/26(80.8%)	184/213(86.4%)	0.462		
			LN metastasis	65/122(53.3%)	89/161(55.3%)	31/52(59.6%)	185/335(55.2%)	0.459	0.441	0.879
			ETE	82/136(60.3%)	111/184(60.3%)	26/55(47.3%)	219/375(58.4%)	0.187	0.618	0.877
3	rs9858271	*FHIT*		AA	A**G**	**GG**				
			BRAF^{V600E}	35/45(77.8%)	101/111(91.1%)	47/56(83.9%)	183/212(86.3%)	0.478		
			LN metastasis	42/74(56.8%)	98/175(56.0%)	42/82(51.2%)	182/331(55.0%)	0.492	0.328	0.424
			ETE	49/81(60.5%)	114/191(59.7%)	55/99(55.6%)	218/371(58.8%)	0.494	0.934	0.105
4	rs1874564	*SEPT11*		AA	A**G**	**GG**				
			BRAF^{V600E}	11/12(91.7%)	68/76(89.5%)	106/126(84.1%)	185/214(86.4%)	0.236		
			LN metastasis	13/20(65.0%)	67/118(56.8%)	106/196(54.1%)	186/334(55.7%)	0.358	0.229	0.381
			ETE	11/22(50.0%)	75/130(57.7%)	134/223(60.1%)	220/375(58.7%)	0.372	0.688	**0.009**
8	rs12542743	*NRG1*		**CC**	**C**T	TT				
			BRAF^{V600E}	23/27(85.2%)	89/103(86.4%)	72/83(86.7%)	184/213(86.4%)	0.855		
			LN metastasis	27/38(71.1%)	83/155(53.5%)	75/140(53.6%)	185/333(55.6%)	0.156	**0.025**	0.545
			ETE	26/43(60.5%)	110/173(63.6%)	82/157(52.2%)	218/373(58.4%)	0.097	0.152	0.752
8	rs6996585	*NRG1*		AA	A**G**	**GG**				
			BRAF^{V600E}	80/92(87.0%)	81/92(88.0%)	24/29(82.8%)	185/213(86.9%)	0.700		
			LN metastasis	85/153(55.6%)	72/146(49.3%)	26/35(74.3%)	183/334(54.8%)	0.348	**0.015**	0.585
			ETE	96/170(56.5%)	97/166(58.4%)	24/37(64.9%)	217/373(58.2%)	0.393	0.067	0.377

Chr	SNP	Gene	Clinical phenotype	Genotypes			Total	*P*-value	*BRAF* positive	*BRAF* negative
8	rs2439302	*NRG1*		CC	C**G**	**GG**				
			BRAF^{V600E}	84/96(87.5%)	71/85(83.5%)	18/21(85.7%)	173/202(85.6%)	0.589		
			LN metastasis	87/159(54.7%)	65/134(48.5%)	19/23(82.6%)	171/316(54.1%)	0.322	**0.022**	0.908
			ETE	99/175(56.6%)	88/153(57.5%)	17/25(68.0%)	204/353(57.8%)	0.430	0.107	0.376
9	rs72753537	*FOXE1*		**CC**	**C**T	TT				
			BRAF^{V600E}	5/6(83.3%)	44/48(91.7%)	136/160(85.0%)	185/214(86.4%)	0.386		
			LN metastasis	6/8(75.0%)	35/68(51.5%)	145/260(55.8%)	186/336(55.4%)	0.907	0.460	0.636
			ETE	4/9(44.4%)	44/77(57.1%)	173/290(59.7%)	221/376(58.8%)	0.392	0.566	0.212
12	rs11175834	*MSRB3*		CC	C**T**	**TT**				
			BRAF^{V600E}	116/132(87.9%)	56/67(83.6%)	10/12(83.3%)	182/211(86.3%)	0.392		
			LN metastasis	123/218(56.4%)	48/95(50.5%)	11/18(61.1%)	182/331(55.0%)	0.691	0.403	0.746
			ETE	145/242(59.9%)	61/108(56.5%)	10/18(55.6%)	216/368(58.7%)	0.506	0.600	0.533
12	rs16934253	*SLC24A6*		**AA**	**A**G	GG				
			BRAF^{V600E}	0/0 (0%)	22/25(88.0%)	164/190(86.3%)	186/215(86.5%)	0.822		
			LN metastasis	0/0 (0%)	17/37(45.9%)	169/300(56.3%)	186/337(55.2%)	0.245	0.064	0.127
			ETE	0/0 (0%)	23/39(59.0%)	198/338(58.6%)	221/377(58.6%)	0.963	0.592	0.392
14	rs34081947	*NKX2-1*		CC	C**T**	**TT**				
			BRAF^{V600E}	59/63(93.7%)	91/107(85.0%)	35/43(81.4%)	185/213(86.9%)	0.056		
			LN metastasis	50/92(54.3%)	94/172(54.7%)	42/72(58.3%)	186/336(55.4%)	0.633	0.509	0.726
			ETE	58/102(56.9%)	119/193(61.7%)	43/80(53.8%)	220/375(58.7%)	0.756	0.945	0.836
14	rs944289	*NKX2-1*		CC	C**T**	**TT**				
			BRAF^{V600E}	52/58(89.7%)	89/104(85.6%)	45/53(84.9%)	186/215(86.5%)	0.450		
			LN metastasis	43/84(51.2%)	91/167(54.5%)	52/86(60.5%)	186/337(55.2%)	0.221	0.206	0.634
			ETE	53/92(57.6%)	112/192(58.3%)	56/93(60.2%)	221/377(58.6%)	0.721	0.338	0.758
19	rs7248104	*INSR*		**AA**	**A**G	GG				
			BRAF^{V600E}	34/39(87.2%)	86/98(87.8%)	60/71(84.5%)	180/208(86.5%)	0.623		
			LN metastasis	30/57(52.6%)	87/164(53.0%)	63/108(58.3%)	180/329(54.7%)	0.419	0.355	0.156
			ETE	49/70(70.0%)	107/181(59.1%)	59/116(50.9%)	215/367(58.6%)	**0.010**	0.571	**0.001**

All analyses were conducted with papillary thyroid cancer. Bold genotype indicates risk allele. Bold *P*-value indicates significance of < 0.05. Chr, chromosome number; ETE, extrathyroidal extension; LN, lymph node; OR, odd ratio; SNP, single nucleotide polymorphism

The most significantly associated variant in the NRG1 locus

The most significant association was identified in the intronic region of *NRG1* at 8p12. In this locus, SNPs on the *NRG1* gene were shown to have a more significant association with PTC than with DTC (Figure 3 and 4, Table 5). In a joint analysis, rs6996585 of *NRG1* was the most significant signal ($P = 9.01 \times 10^{-12}$ in PTC, 1.08×10^{-10} in DTC). We found a significant *cis*-eQTL of rs6996585 for *NRG1* expression in the thyroid tumour tissue ($P = 0.0053$, Figure 5a, Table 6). A similar expression pattern was shown in the normal thyroid tissue, although it was not statistically significant (P-value = 0.0526). However, the predicted expression result of discovery case samples that used normal thyroid reference data, showed a highly significant association ($P = 2.99 \times 10^{-244}$, Figure 5a, Table 6). The public expression data also showed a significant association ($P = 5.79 \times 10^{-21}$) in the normal thyroid tissue (Figure. 6a, and Table 7). Furthermore, the GSEA result indicated that rs6996585 was significantly associated with 31 gene set pathways related to cellular growth signals or cancer in the normal thyroid tissue (FDR $q < 0.05$, Tables 9 and 10). We confirmed that the common genes from 31 significant gene sets were enriched in the *ERBB-MAPK* signalling pathway (Figure. 7a). A clinical phenotype analysis showed that rs6996585 was associated with lymph node metastasis in patients with $BRAF^{V600E}$ mutated tumours ($P = 0.015$, Figure 7b

and Table 11).

In this region, another candidate SNP (rs12542743) had a similar association (Joint $P = 1.01 \times 10^{-10}$ in PTC, 4.61×10^{-10} in DTC, Figure. 4a and 4b) and *cis*-eQTL results (Figure 5b, Table 6). Although the previously reported SNP (rs2439302) showed a marginal association ($P = 1.50 \times 10^{-6}$ in PTC, 8.38×10^{-5} in DTC) in the discovery stage, the *cis*-eQTL of this SNP was more significantly associated than that of rs6996585 (Figure 5c and 6c). These two SNPs also showed a similar association with lymph node metastasis in the $BRAF^{V600E}$ mutation-positive group (Table 11).

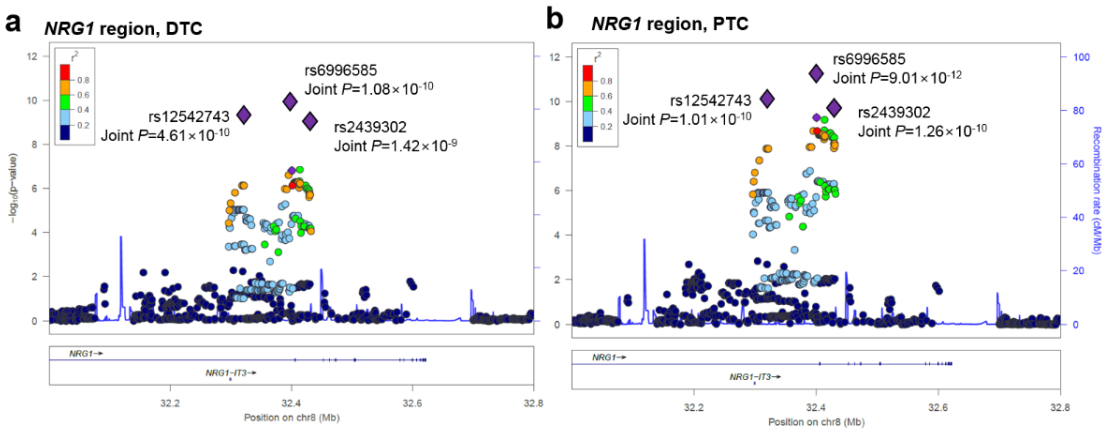

Figure 4. Regional association plots of the most associated variant in *NRG1* locus. A regional association plot for (a) DTC and (b) PTC. The purple large diamonds indicate the joint analyses of associated SNP and nearby SNPs are color coded according to the level of LD with the top SNP. The left y-axis shows the significance of the association with $-\log_{10}$ scale, and the right y-axis shows a recombination rate across the region. Estimated recombination rates from the 1000 Genome ASN, hg19 database are plotted with the blue line to reflect the local LD structure.

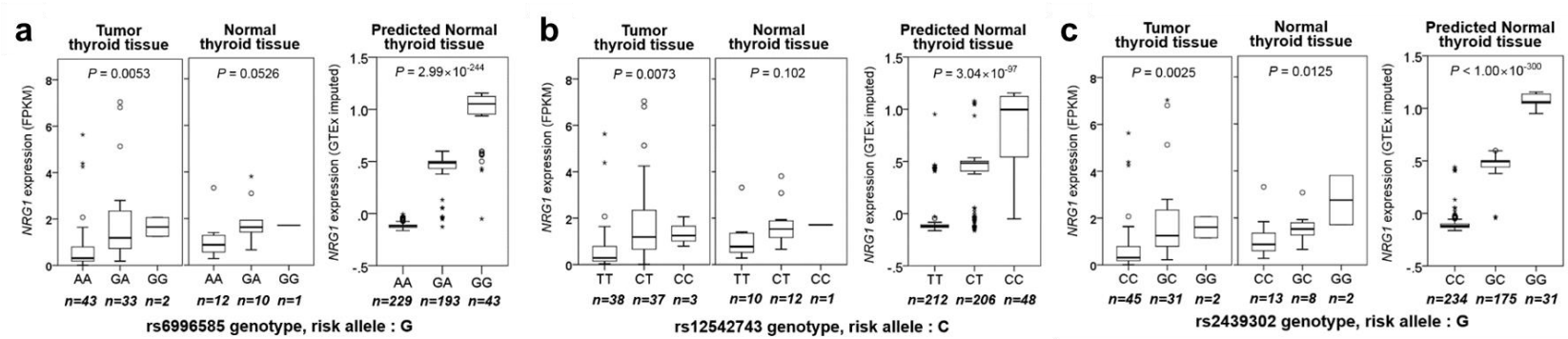

Figure 5. Expression of the most associated variant in *NRG1* locus.

The *cis*-eQTL result of *NRG1* in tumour and normal thyroid tissues and predicted expression of 470 DTC cases according to the (a) rs6996585, (b) rs12542743 and (c) rs2439302 genotypes.

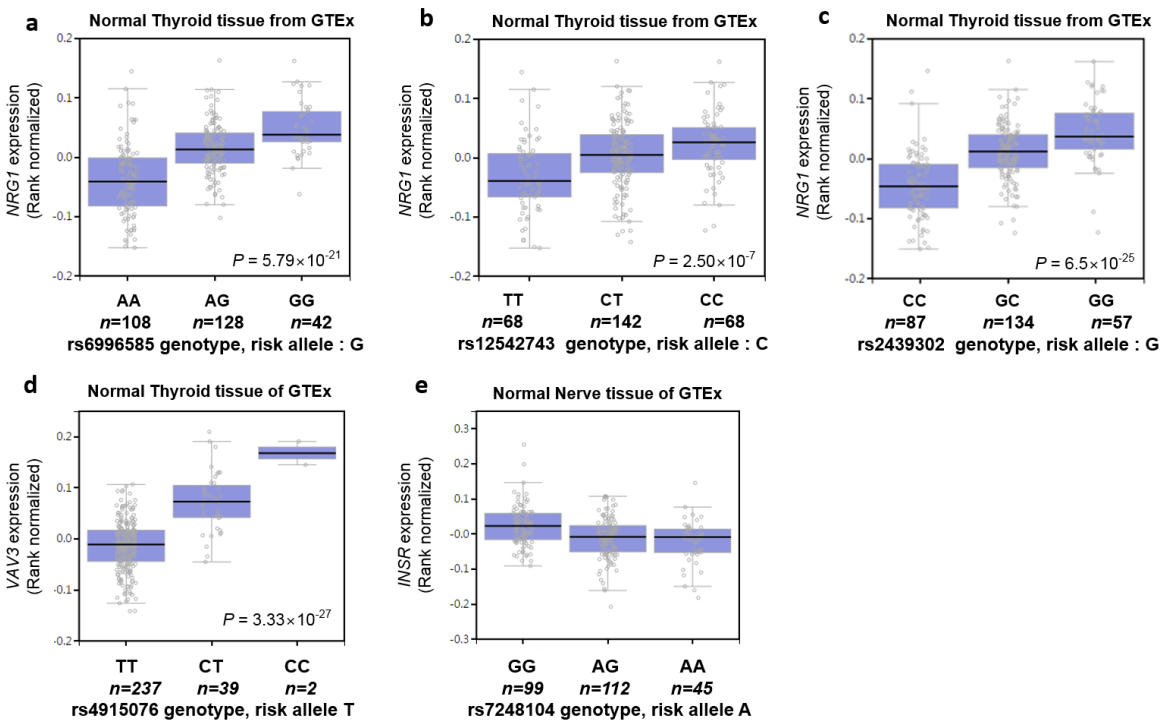

Figure 6. *Cis*-eQTL result from GTEx public data. The *cis*-eQTL result of *NRG1* in normal thyroid tissue according to the (a) rs6996585, (b) rs12542743 and (c) rs2439302 genotypes. The *cis*-eQTL result of (d) *VAV3* in normal thyroid tissue according to the rs4915076 genotype and (e) *INSR* in normal nerve tissue according to the rs7248104 genotype from GTEx2015 v6 data (http://www.gtexportal.org).

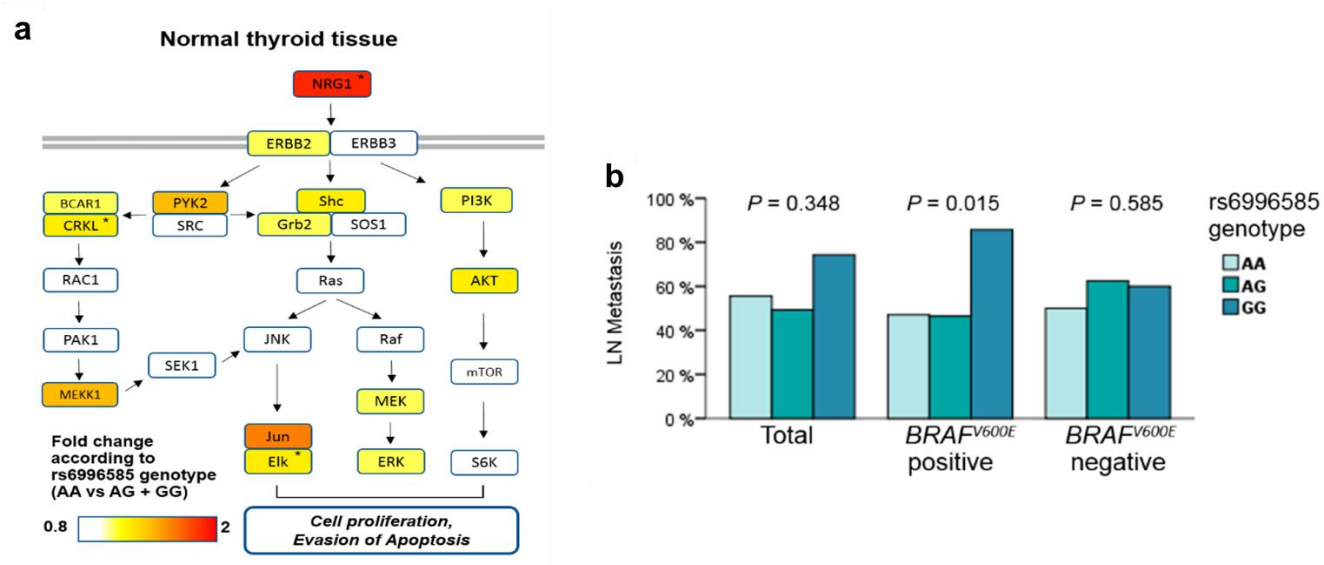

Figure 7. Gene Set enrichment analysis of the most associated variant in *NRG1* locus.

(a) The characteristic gene expression of normal thyroid tissues. The genes in *ERBB-MAPK* signaling pathway were represented by fold change according to rs6996585 genotype (AA vs AG + GG). Asterisk indicates significant fold change with *P*-value < 0.05. (b) Lymph node metastasis according to the rs6996585 genotypes and *BRAF* mutation.

Other known associated variants in the NKX2-1, DIRC3, or FOXE1 loci

At 14q13.3 in a region near *NKX2-1*, two SNPs (rs34081947 and rs944289) were significantly associated with DTC (Joint $P = 1.19 \times 10^{-7}$ and 1.39×10^{-6}, respectively, Figure 8a). The top-ranked variant (rs34081947) in this region showed a significant association ($P = 0.0323$) with the *NKX2-1* expression level in the tumour tissue (Figure. 8d). The rs944289 showed a greater significant association with *NKX2-1* and *SFTA3* expression levels compared to rs34081947 in both the tumour and normal tissues (Figure. 8e and f, Table 6).

At 2q35, rs12990503 in the intron of *DIRC3* gene was significantly associated (Joint $P = 3.55 \times 10^{-9}$) with DTC (Figure 8b). In the tumour tissue, this SNP showed a *cis*-eQTL for *TNS1* (Figure 8g). We could not find a *cis*-eQTL of *DIRC3* in our data, but the public expression data showed its *cis*-eQTL ($P = 7.19 \times 10^{-6}$) in skin tissue (Figure. 8h, Table 8). The GSEA result demonstrated that this SNP was associated with *TGF-beta* receptor signalling in the thyroid tumour tissue (Table 9).

At 9q22.33 near the *FOXE1* region, rs965513 showed the most significant association in the discovery stage of DTC, but it was not replicated. Among the seven candidate SNPs proposed previously (Table 12), rs72753537 was positively replicated and showed a suggestive association (Joint $P = 7.67 \times 10^{-6}$, Figure 8c). This SNP did not have any *cis*-eQTL for nearby genes. However,

the predicted expression result showed a *cis*-eQTL with *C9orf156, CORO2A, XPA* and *TSTD2* genes (Table 6) and the public expression data showed a *cis*-eQTL with *C9orf156* in various tissues (Figure 8i, Table 8). The association of the seven SNPs in our DTC or PTC subjects are summarized in Table 12.

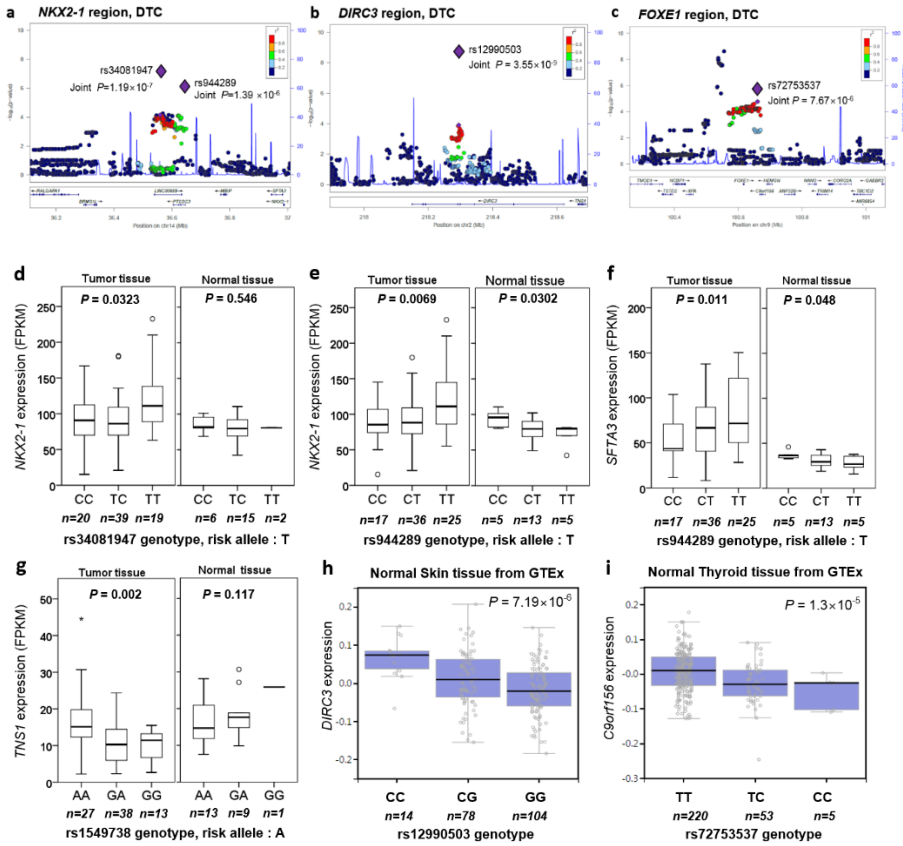

Figure 8. Regional association plots and expression for previously documented loci. A regional association plot for (a) *NKX2-1,* (b) *DIRC3* and (c) *FOXE1*. The purple large diamonds indicate the joint analyses of most or second associated SNP and nearby SNPs are color coded according to the level of LD with the top SNP. The left y axis shows the significance of the association with $-\log_{10}$ scale, and the right y-axis shows a recombination rate across the region. Estimated recombination rates from the 1000 Genome ASN, hg19 database are plotted with the blue line to reflect the local LD structure. The *cis*-eQTL result of *NKX2-1* according to the (d) rs34081947 and (e) rs944289 genotypes, (f) *SFTA3* according to the rs944289 genotypes and (g) *TNS1* according to the rs1549738 genotypes in tumor and normal thyroid tissues. The *cis*-eQTL result of (h) *DIRC3* in skin tissue and (i) *C9orf156* in thyroid tissue from GTEx2015 v6 data (http://www.gtexportal.org) according to the rs12990503 and rs72753537 genotypes, respectively.

Table 12. SNPs of *FOXE1* region and DTC and PTC association in Korean Population

Chr	SNP	Position	Gene	Risk Allele	Stage	DTC			PTC		
						Allele Frequency (case/control)	OR	*P*	Allele Frequency (case/control)	OR	*P*
9	rs965513	100556109	*FOXE1*	A	Discovery	0.11/0.06	1.91	2.35E-09	0.11/0.06	1.94	8.20E-09
					Replication	0.09/0.07	1.25	0.1372	0.09/0.07	1.23	0.1653
					Joint	0.10/0.06	1.67	6.20E-11	0.10/0.06	1.66	4.48E-10
9	rs1588635	100537802	*FOXE1*	A	Discovery	0.11/0.07	1.81	1.74E-08	0.12/0.07	1.83	6.76E-08
					Replication	0.09/0.08	1.18	0.2425	0.09/0.08	1.17	0.2706
					Joint	0.10/0.07	1.58	2.36E-09	0.10/0.07	1.57	1.30E-08
9	rs7028661	100538470	*FOXE1*	A	Discovery	0.11/0.07	1.80	2.52E-08	0.12/0.07	1.82	9.47E-08
					Replication	0.09/0.08	1.20	0.2090	0.09/0.08	1.19	0.2343
					Joint	0.10/0.07	1.57	3.08E-09	0.10/0.07	1.56	1.64E-08
9	rs10122541	100628268	*FOXE1*	G	Discovery	0.12/0.08	1.57	1.43E-05	0.10/0.06	1.81	4.07E-07
					Replication	0.10/0.08	1.22	0.1690	0.08/0.07	1.03	0.8341
					Joint	0.11/0.08	1.36	4.25E-06	0.09/0.06	1.48	4.77E-06
9	rs72753537	100660746	*FOXE1*	C	Discovery	0.12/0.07	1.63	3.56E-06	0.12/0.07	1.76	1.70E-07
					Replication	0.09/0.07	1.38	**0.0352**	0.09/0.07	1.43	**0.0209**
					Joint	0.10/0.07	1.41	7.67E-06	0.11/0.07	1.48	5.37E-07
9	rs7037324	100658318	*FOXE1*	A	Discovery	0.12/0.08	1.56	1.70E-05	0.19/0.13	1.56	1.43E-06
					Replication	0.10/0.08	1.28	0.0865	0.15/0.14	1.11	0.3758
					Joint	0.11/0.08	1.38	1.28E-05	0.17/0.13	1.32	1.62E-05
9	rs1867277	100615914	*FOXE1*	A	Discovery	0.17/0.12	1.44	4.38E-05	0.18/0.12	1.54	4.96E-06
					Replication	0.12/0.13	0.89	0.3109	0.11/0.13	0.86	0.1982
					Joint	0.14/0.12	1.18	0.0176	0.15/0.12	1.19	0.0146

The SNP positions are indexed to the National Center for Biotechnology Information (NCBI) build 37. Bold indicates significance of $P < 0.05$ in replication stage.
Chr, chromosome number; OR, odd ratio; SNP, single nucleotide polymorphism.

Novel candidate variants in the VAV3, PCNXL2, INSR, MRSB3, FHIT or SEPT11 loci

Using the expression data, we evaluated the six newly discovered regions. At 1p13.3, an intronic region of *VAV3*, rs4915076 was significantly associated with PTC (Joint $P = 7.09 \times 10^{-8}$, Figure 9a) and is a *cis*-eQTL for *VAV3* expression in the tumour tissue ($P = 0.017$, Figure 9g). The predicted expression result showed a highly significant association ($P < 1.00 \times 10^{-300}$, Table 6) and the public expression data also showed a significant association ($P = 3.33 \times 10^{-27}$) in thyroid tissue (Figure. 6d, Table 7). The GSEA result showed that rs4915076 was associated with the steroid hormone pathways in the tumour thyroid tissue (Table 9). In the intron of *PCNXL2* at 1q42.2, rs4649295 showed a significant association with DTC (Joint $P = 6.00 \times 10^{-8}$, Figure 9b) and is a *cis*-eQTL for *PCNXL2* expression in the tumour tissue ($P = 0.003$, Figure 9h). Rs9858271, near *FHIT* at 3p14.2, showed a significant association with PTC (Joint $P = 2.76 \times 10^{-8}$, Figure 9c). Rs1874564, near *SEPT11* at 4q21.1, showed a significant association with DTC (Joint $P = 2.04 \times 10^{-7}$, Figure 9d). On chromosome 12q14.3, a SNP (rs11175834, Joint $P = 4.26 \times 10^{-8}$) significantly associated with DTC was located near *MSRB3* (Figure 9e). In the intronic region of the *INSR* gene at 19p13.2, rs7248104 was positively replicated and showed a

suggestive association with DTC (Joint $P = 2.00 \times 10^{-5}$, Figure 9f). However, we could not find any *cis*-regulation of nearby gene expression in the *FHIT, SEPT11, MSRB3* and *INSR* regions with our expression data, but the predicted expression result showed a *cis*-eQTL for *INSR* ($P = 4.91 \times 10^{-41}$, Table 6) and the public expression data showed a *cis*-eQTL for *INSR* in nerve tissue and for *SEPT11* in whole blood (Table 8). The GSEA result showed that a SNP of *INSR* was associated with the *ERK* pathway, and a SNP of *SEPT11* was associated with the *ATM* pathway (Table 9). The association analysis of the clinical phenotypes showed that the SNPs of *SEPT11* and *INSR* were associated with extrathyroidal extension in the $BRAF^{V600E}$ mutation-negative group (Table 11).

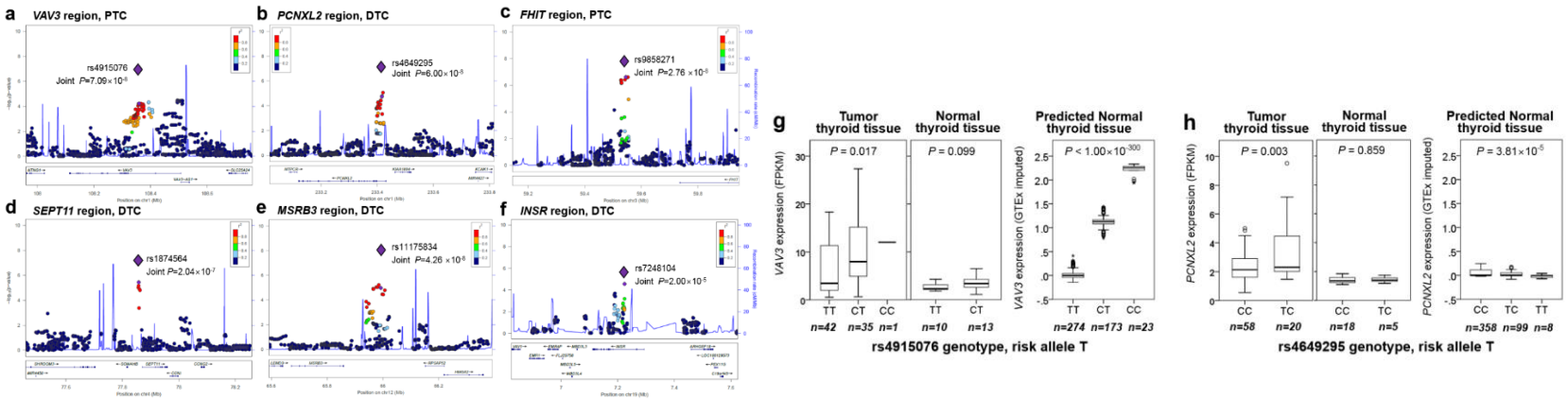

Figure 9. Regional association plots and expression of the novel DTC associated variant in *VAV3*, *PCNXL2, FHIT, SEPT11, MSRB3, and INSR* locus. A regional association plot for (a) *VAV3*, (b) *PCNXL2*, (c) *FHIT*, (d) *SEPT11*, (e) *MSRB3* and (f) *INSR* regions. The purple large diamonds indicate the joint analyses of most associated SNP and nearby SNPs are color coded according to the level of LD with the top SNP. The left y-axis shows the significance of the association with $-\log_{10}$ scale, and the right y-axis shows a recombination rate across the region. Estimated recombination rates from the 1000 Genome ASN, hg19 database are plotted with the blue line to reflect the local LD structure. The *cis*-eQTL result of (g) *VAV3* and (h) *PCNXL2* in tumour and normal thyroid tissues and predicted expression of 470 DTC cases according to the rs4915076 and rs4649295 genotypes, respectively.

A comparison with the European GWAS results

The effect sizes (OR) and *P*-values reported in the previous GWAS in Europeans (Table 1) were compared with those from our Stage 1 genome scan result (Figure 10, Table 13). Fourteen SNPs previously reported in Europeans were available for analyses in our study. Four SNPs of *FOXE1* and one SNP of *NRG1* showed a similar effect size and a similar significant association of $P < 0.0001$. The SNPs of *NKX2-1* and *DIRC3* showed a similar or lesser effect size and a nominal association of $P < 0.05$. The other six SNPs of *IMMP2L, DHX35, ARSB* and *WDR11-AS1* showed no association in a Korean population.

In addition, the novel associated SNPs of this study were compared with previously reported SNPs. As for DTC, the SNPs of *NRG1* showed a lower association than those of *FOXE1*. However, as for PTC, rs6996585 of *NRG1* showed the most significant *P*-value of the previous reported SNPs. Furthermore, the novel associated regions (*VAV3, PCNXL2, INSR, MRSB3, FHIT* or *SEPT11*) were located in the middle of previously reported SNPs.

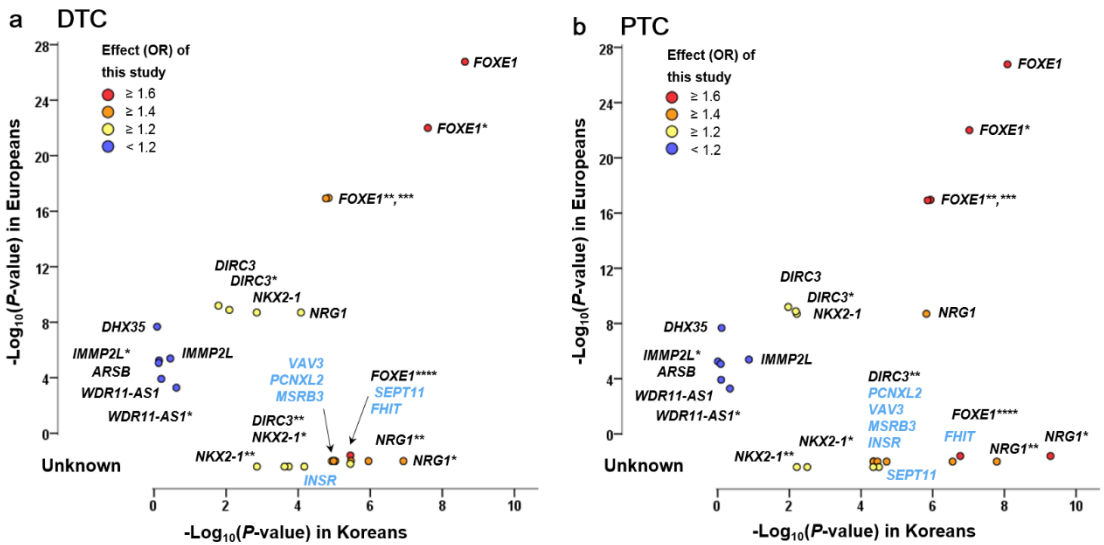

Figure 10. A comparison of association result for (a) DTC or (b) PTC between Europeans and Koreans. The *P*-values for Koreans (x-axis) and Europeans (y-axis) are plotted with the corresponding Korean effect sizes (OR; Odd Ratio). The *P*-value shows the $-\log_{10}$ scale and the *P*-value of novel SNPs from this study are compared as unknown. The novel gene of this study are shown in blue. The SNPs of same gene were distinguished by *FOXE1* for rs965513, *FOXE1** for rs7028661, *FOXE1**** for rs10122541, *FOXE1***** for rs72753537, *DIRC3* for rs966423, *DIRC3** for rs6759952, *NRG1* for rs2439302, *NRG1** for rs6996585, *NRG1*** for rs12542743, *NKX2-1* for rs944289, *NKX2-1** for rs34081947, *NKX2-1*** for rs944289, *DIRC3* for rs6759952, *DIRC3** for rs966423, *DIRC3*** for rs12990503, *IMMP2L* for rs10238549, *IMMP2L** for s7800391, *WDR11-AS1* for rs2997312 and *WDR11-AS1** for rs10788123.

Table 13. A comparison of previous reported SNPs for associated with DTC or PTC between Europeans and Koreans.

Gene	SNP	Europeans			DTC in Koreans		PTC in Koreans	
		OR	*P*-value	population	OR	*P*-value	OR	*P*-value
FOXE1	rs965513	1.75	1.7×10^{-27}	Iceland etc.	1.91	2.4×10^{-9}	1.94	8.2×10^{-9}
FOXE1	rs7028661	1.64	1.0×10^{-22}	Spain	1.80	2.5×10^{-8}	1.82	9.5×10^{-8}
FOXE1	rs10122541	1.54	1.1×10^{-17}	Spain	1.57	1.4×10^{-5}	1.69	1.2×10^{-6}
FOXE1	rs7037324	1.54	1.2×10^{-17}	Spain	1.56	1.7×10^{-5}	1.68	1.4×10^{-6}
NRG1	rs2439302	1.36	2.0×10^{-9}	Iceland etc.	1.34	8.4×10^{-5}	1.48	1.5×10^{-6}
NKX2-1	rs944289	1.37	2.0×10^{-9}	Iceland etc.	1.24	0.0014	1.22	0.0062
DIRC3	rs966423	1.34	1.3×10^{-9}	Iceland etc.	1.24	0.0081	1.27	0.0067
DIRC3	rs6759952	1.21	6.4×10^{-10}	Italy etc.	1.21	0.0164	1.25	0.0107
IMMP2L	rs10238549	1.27	4.1×10^{-6}	Italy etc.	1.10	0.3542	1.17	0.1343
IMMP2L	rs7800391	1.25	5.7×10^{-6}	Italy etc.	1.03	0.7271	1.00	0.9850
DHX35	rs7267944	1.39	2.1×10^{-8}	Italy etc.	0.98	0.8209	0.98	0.7752
ARSB	rs13184587	1.28	8.5×10^{-6}	Italy etc.	1.03	0.7453	1.02	0.8216
WDR11-AS1	rs2997312	1.35	1.2×10^{-4}	Spain	0.94	0.6299	0.97	0.8039
WDR11-AS1	rs10788123	1.26	5.2×10^{-4}	Spain	0.92	0.2401	0.94	0.4540

OR, odd ratio; SNP, single nucleotide polymorphism.

Table 14. Variants of recently reported six novel and replicating loci for thyroid cancer in Korean Population.

Locus	SNP	Position	Annotation / Nearby gene(s)	P-value of Gudmundsson et al.	Discovery result in this study		
					P-value	Nearby top SNP with P < 0.001	P-value (Joint P) of nearby top SNP
1q42.2	rs12129938	233,276,815	Intron variant *PCNXL2*	4.0E-11	0.002	rs4649295	1.04E-05
3q26.2	rs6793295	169,800,667	Missense variant *TERC, LRRC34*	2.7E-8	0.0474	-	
5p15.33	rs10069690	1,279,675	Intron variant *TERT*	3.2E-7	NA	-	
5q22.1	rs73227498	112,150,207	Intergenic variant *NREP, EPB41L4A*	3.0E-10	NA	-	
10q24.33	rs7902587	103,934,543	Intergenic variant *OBFC1*	5.4E-11	NA	rs4244255	8.72E-06
15q22.33	rs2289261	67,165,147	Intron variant *SMAD3*	3.1E-9	0.2844	-	

Chapter II. Genome-wide association and expression quantitative trait loci studies for thyroid nodule

Methods

Discovery series and thyroid ultrasonography

A discovery genome-wide association scan was conducted in the Ansung cohort, a community-based cohort living in rural Korea. The Ansung cohort is one of the two community-based cohorts of the Korean Association Resource (KARE) project, and the detailed information regarding the KARE project has been described elsewhere (43). In 5,018 cohort participants, 3,161 individuals were evaluated by thyroid ultrasonography between 2011 and 2012. Subsequent thyroid ultrasonography was performed in 2,561 of the 3,161 participants between 2013 and 2014. From 2,561 subjects who underwent two consecutive ultrasonographies, 952 showed solid nodules and 824 showed normal thyroid glands, persistently in repeated sonographic images (Figure 11). The remaining 785 individuals had pure cystic nodules, disappeared nodules, newly appeared nodules, sonographic evidence of thyroidectomy or history of thyroid cancer. The disappeared or newly developed nodules were not true solid nodules but were presumed as parts of focal or diffuse thyroiditis. Individuals with pure cystic nodules were also not included. The characteristics of the participants are described in Table 15.

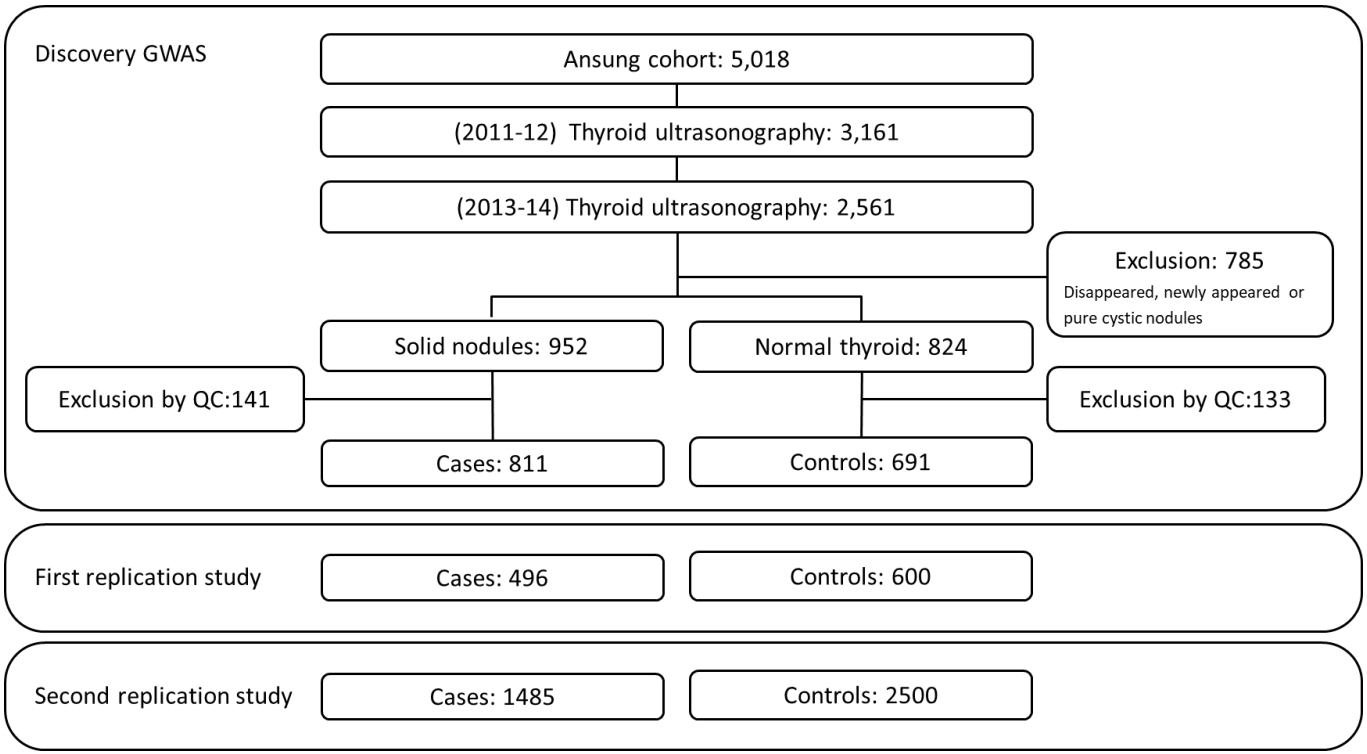

Figure 11. Subjects of the discovery stage GWAS and the replication studies for thyroid nodules

Table 15. Descriptive characteristics of the participants.

Characteristics	Total (joint)	Discovery	First replication	Second replication
Thyroid nodule				
Number	2792	811	496	1485
Age, years ± SD	54 ± 1	56 ± 8	52 ± 8	53 ± 1
Male (%)	1210 (43.3%)	237 (29.2%)	224 (45.2%)	749 (50.4%)
Normal thyroid				
Number	3791	691	600	2500
Age, years ± SD	51 ± 1	53 ± 8	59 ± 4	49 ± 1
Male (%)	2557 (67.5%)	382 (55.3%)	479 (79.8%)	1696 (67.8%)
Thyroid cancer				
Number		470		
Age, years ± SD		44 ± 13		
Male (%)		61(13.0%)		

First replication series and ultrasonography

For a replication study, we used 496 cases that had benign solid nodules without features of diffuse or focal thyroiditis, and 600 controls showing normal thyroid glands in ultrasonography and normal thyroid function in a thyroid hormone study. All cases and controls were selected from the participants of the Cancer Screening Program at the National Cancer Center, Republic of Korea. All subjects with nodules were not previously diagnosed with thyroid cancer. The characteristics of the subjects are shown in Table 15.

Second replication

For the second replication study, we requested data from the Gene-environmental interaction and phenotype (GENIE) cohort, a sub-cohort of The Health and Prevention Enhancement (H-PEACE), which is a retrospective, population-based cohort study conducted at the Seoul National University Hospital Gangnam Center in Korea. Detailed information about GENIE cohort has been described in a previous report (51-53). Among 6,579 subjects who were analyzed genome-wide SNP arrays, 5,737 had reports of thyroid ultrasonography in electronic health record. Results of thyroid ultrasonography showed that 1,485 had solid thyroid nodules and 2,500 had normal thyroid

glands. The remaining 1,752 subjects had pure cysts, nodules with ultrasonographic features of highly suspicious of malignancy, diffuse or focal thyroiditis, operated thyroid gland or history of thyroid cancer. The characteristics of cases and controls included in second replication study are shown in Table 15.

Discovery GWAS and Imputation

Genotyping for KARE projects has been described previously (43). Briefly, peripheral blood samples were obtained from the participants and genomic DNA was extracted from the lymphocytes. Genotyping for discovery series was performed using the Affymetrix Genome-Wide Human SNP array 5.0 (Affymetrix, Santa Clara, CA, USA). For quality control, individuals with low call rates (< 96%), high heterozygosity (> 30%), or gender inconsistencies were removed, and subjects with genetic relatedness were excluded using the kinship identical-by-descent (IBD) coefficient (Z0 > 0.8). In total, 141 cases and 133 controls were excluded by individual quality control procedures. SNPs with a minor allele frequency of < 1%, SNPs with a missing genotype rate ≥ 5%, and SNPs whose genotype frequencies departed from the Hardy-Weinberg equilibrium at $P < 1 \times 10^{-6}$ were excluded. Imputation was performed using the

IMPUTE2 software(http://mathgen.stats.ox.ac.uk/impute/impute_v2.html) and the 1000 genome ASN Phase I panel (NCBI build37, hg19) was used as a reference. For imputation quality control, SNPs with minor allele frequency (MAF) < 0.01 or SNP missing rate > 0.05 were removed, and the remaining imputed SNPs were combined with the original KARE SNPs. Finally, 811 cases and 691 controls were analyzed for a total of 3,996,558 markers.

Candidate SNP and genotyping of first replication

From the genome-wide association analysis, the best 1,000 SNPs ($P < 3 \times 10^{-4}$) were screened and 27 representative SNPs in 26 distinct genomic regions were selected as candidate markers. We also included 2 SNPs (rs944289 on *MBIP/NKX2-1* and rs11175834 on *MSRB3*) previously reported as signals for thyroid cancer (Table 16). For the replication study, genotyping was performed using the Fluidigm SNP Type Assay platform (Fluidigm, South San Francisco, USA) for candidate SNPs. All SNPs were successfully genotyped with call rates > 95% and Hardy-Weinberg equilibrium P > 0.001.

Table 16. Two-stage of GWAS for thyroid nodule

Chr SNP Position	Nearest Genes Minor/major allele	Series	Minor allele frequency Case/Control	OR (95% CI)	P-value
1q43	Intron *RYR2*	Discovery	0.40/0.47	0.73(0.62-0.85)	3.74×10^{-5}
rs10925394	G/A	Replication	0.45/0.44	1.09(0.89-1.35)	0.371
237545098		Joint	0.42/0.46	0.84(0.75-0.95)	3.68×10^{-3}
2q23.3	3.3kb 3' of *NMI*	Discovery	0.04/0.08	0.52(0.38-0.72)	7.51×10^{-5}
rs16829835	C/A	Replication	0.07/0.07	1.18(0.79-1.77)	0.404
152123703		Joint	0.06/0.07	0.71(0.56-0.90)	5.00×10^{-3}
2q33.3	Intron *PARD3B*	Discovery	0.45/0.53	0.72(0.62-0.85)	4.70×10^{-5}
rs6721320	C/T	Replication	0.47/0.48	1.04(0.84-1.28)	0.743
206412487		Joint	0.46/0.51	0.82(0.73-0.92)	7.76×10^{-4}
3p25.3	Intron *SLC6A11*	Discovery	0.19/0.25	0.69(0.58-0.83)	6.09×10^{-5}
rs9826851	C/G	Replication	0.29/0.33	0.97(0.81-1.15)	0.696
10896848		Joint	0.23/0.29	0.81(0.73-0.91)	4.83×10^{-4}
3q25.32	Intron *SCHIP1*	Discovery	0.47/0.41	1.39(1.19-1.63)	2.70×10^{-5}
rs4680504	C/G	Replication	0.46/0.46	1.04(0.84-1.27)	0.748
159140563		Joint	0.47/0.43	1.18(1.05-1.33)	4.40×10^{-3}
5p15.32	518kb 5' of *LOC340094*	Discovery	0.12/0.18	0.65(0.53-0.81)	7.83×10^{-5}
rs11134055	T/C	Replication	0.16/0.17	0.92(0.69-1.21)	0.529
4516670		Joint	0.14/0.17	0.76(0.65-0.89)	6.52×10^{-4}
5p13.3	425kb 3' of *LOC729862*	Discovery	0.18/0.13	1.55(1.25-1.93)	8.00×10^{-5}
rs1864034	C/G	Replication	0.16/0.14	1.22(0.93-1.62)	0.156
29352194		Joint	0.17/0.14	1.37(1.17-1.61)	1.03×10^{-4}
5q35.1	17kb 5' of *DUSP1*	Discovery	0.35/0.40	0.71(0.61-0.84)	4.09×10^{-5}
rs17075176	C/T	Replication	0.37/0.39	0.87(0.70-1.09)	0.231
172215412		Joint	0.36/0.40	0.80(0.71-0.90)	3.56×10^{-4}
6p22.1	Intron *ZNF193*	Discovery	0.06/0.11	0.54(0.41-0.71)	1.02×10^{-5}
rs16893827	C/T	Replication	0.08/0.06	1.29(0.86-1.93)	0.219
28197469		Joint	0.07/0.09	0.72(0.58-0.89)	2.80×10^{-3}
6q16.3	967kb 3' of *GRIK2*	Discovery	0.18/0.24	0.67(0.55-0.81)	3.64×10^{-5}
rs4543404	T/A	Replication	0.23/0.22	1.14(0.89-1.46)	0.318
103485033		Joint	0.20/0.23	0.83(0.72-0.96)	1.08×10^{-2}
6q22.31	164kb 5' of *LOC643623*	Discovery	0.04/0.08	0.52(0.38-0.72)	7.80×10^{-5}
rs6939104	A/T	Replication	0.06/0.06	1.03(0.67-1.57)	0.903
125831838		Joint	0.05/0.07	0.71(0.55-0.90)	4.78×10^{-3}
7p15.3	Intron *DFNA5*	Discovery	0.14/0.19	0.65(0.53-0.8)	5.06×10^{-5}
rs2521769	G/A	Replication	0.16/0.16	1.00(0.76-1.32)	0.997
24784485		Joint	0.15/0.18	0.77(0.66-0.91)	1.51×10^{-3}
7q11.22	816kb 5' of *AUTS2*	Discovery	0.10/0.06	1.79(1.34-2.4)	8.02×10^{-5}
rs2711481	A/G	Replication	0.09/0.09	1.23(0.86-1.75)	0.258
68247814		Joint	0.09/0.07	1.33(1.08-1.64)	7.17×10^{-3}
7q31.31	Intron *KCND2*	Discovery	0.43/0.36	1.37(1.17-1.6)	1.12×10^{-4}
rs1860705	T/C	Replication	0.38/0.38	0.97(0.78-1.21)	0.772

120056669		Joint	0.41/0.37	1.22(1.09-1.38)	9.84×10^{-4}
8p23.2	342kb 5' of *CSMD1*	Discovery	0.28/0.35	0.68(0.58-0.81)	6.42×10^{-6}
rs10282750	G/C	Replication	0.33/0.34	1.03(0.83-1.28)	0.800
5194473		Joint	0.30/0.34	0.80(0.71-0.90)	3.83×10^{-4}
8p23.1	55kb 5' of *LOC157273*	Discovery	0.03/0.06	0.45(0.31-0.66)	4.62×10^{-5}
rs330010	T/G	Replication	0.06/0.07	0.98(0.64-1.48)	0.913
9127516		Joint	0.04/0.07	0.66(0.50-0.85)	1.58×10^{-3}
9q21.13	Intron *TRPM3*	Discovery	0.27/0.21	1.42(1.19-1.7)	7.90×10^{-5}
rs4745021	T/A	Replication	0.26/0.22	1.26(0.99-1.60)	0.065
73267607		Joint	0.27/0.22	1.35(1.18-1.55)	1.03×10^{-5}
9q22.33	Intron *CORO2A*	Discovery	0.34/0.41	0.73(0.62-0.86)	1.07×10^{-4}
rs4743176	A/C	Replication	0.36/0.37	0.95(0.76-1.18)	0.615
100930659		Joint	0.35/0.39	0.80(0.71-0.90)	2.79×10^{-4}
10p11.22	Intron *CCDC7*	Discovery	0.12/0.18	0.63(0.51-0.78)	2.32×10^{-5}
rs10508774	G/A	Replication	0.13/0.15	1.05(0.79-1.41)	0.730
32807861		Joint	0.12/0.16	0.75(0.64-0.88)	5.24×10^{-4}
10p11.22	Intron *ITGB1*	Discovery	0.12/0.18	0.61(0.49-0.75)	5.54×10^{-6}
rs3780873	A/G	Replication	0.14/0.16	1.01(0.76-1.35)	0.927
33213680		Joint	0.13/0.17	0.73(0.62-0.86)	1.33×10^{-4}
10q23.1	63kb 3' of *SH2D4B*	Discovery	0.38/0.46	0.75(0.64-0.87)	1.70×10^{-4}
rs7086866	G/T	Replication	0.38/0.40	0.95(0.77-1.18)	0.655
82469730		Joint	0.38/0.43	0.81(0.72-0.91)	4.86×10^{-4}
10q26.3	98kb 3' of *FLJ46300*	Discovery	0.33/0.39	0.74(0.63-0.86)	1.49×10^{-4}
rs10872824	A/G	Replication	0.38/0.36	0.99(0.80-1.22)	0.909
133506848		Joint	0.35/0.38	0.86(0.77-0.97)	1.48×10^{-2}
12q14.3	132kb 3' of *MSRB3*	Discovery	0.18/0.12	1.58(1.27-1.96)	4.69×10^{-5}
rs11175834	T/C	Replication	0.16/0.14	1.11(0.83-1.49)	0.479
65992636		Joint	0.17/0.13	1.35(1.14-1.59)	3.80×10^{-4}
14q13.3	158kb 3' of *MBIP*	Discovery	0.49/0.42	1.4(1.2-1.64)	1.45×10^{-5}
rs2415317	A/G	Replication	0.45/0.42	1.21(0.97-1.50)	0.087
36609678		Joint	0.47/0.42	1.30(1.15-1.46)	1.58×10^{-5}
14q13.3	119kb 3' of *MBIP*	Discovery	0.49/0.43	1.37(1.18-1.6)	5.72×10^{-5}
rs944289	T/C	Replication	0.46/0.43	1.20(0.97-1.48)	0.095
36649246		Joint	0.48/0.43	1.27(1.13-1.43)	7.53×10^{-5}
14q23.3	5kb 3' of *AKAP5*	Discovery	0.51/0.44	1.35(1.16-1.57)	1.35×10^{-4}
rs1742159	G/C	Replication	0.49/0.49	0.94(0.77-1.15)	0.540
64946191		Joint	0.51/0.48	1.18(1.06-1.33)	4.10×10^{-3}
15q12	Intron *GABRB3*	Discovery	0.14/0.20	0.66(0.54-0.82)	9.59×10^{-5}
rs1426223	T/C	Replication	0.19/0.18	1.10(0.84-1.43)	0.498
26952294		Joint	0.16/0.19	0.81(0.69-0.94)	5.52×10^{-3}
18p11.31	Intron *EPB41L3*	Discovery	0.19/0.27	0.61(0.5-0.73)	1.83×10^{-7}
rs9952940	C/T	Replication	0.21/0.24	0.77(0.60-0.99)	0.045
5472021		Joint	0.20/0.26	0.69(0.60-0.80)	4.69×10^{-7}
19p13.2	Intron *FBN3*	Discovery	0.18/0.12	1.61(1.3-2)	1.67×10^{-5}
rs17261689	C/T	Replication	0.15/0.14	1.23(0.91-1.66)	0.171
8136027		Joint	0.17/0.13	1.43(1.22-1.69)	1.79×10^{-5}

Chr, Chromosome; CI, confidence interval; OR, odds ratio; SNP. Single-nucleotide polymorphism
The SNP positions are indexed to the National Center for Biotechnology Information (NCBI) build 37.

Genotyping of second replication

Subjects of GENIE cohort were genotyped by DNAIink, Inc. using Affymetrix Axiom™ KORV1.1-96 Genotyping Arrays (Affymetrix, Santa Clara, CA, USA). More information on genotyping, quality control procedures and imputation can be found in previous papers(52, 53). We obtained genotypes in cases and controls for four SNPs that were selected from discovery GWAS and first replication study.

Comparison of allele frequencies between DTC, thyroid nodules, and normal thyroid

We compared the allele frequencies for SNPs identified from the GWAS for DTC (Table 5) and the present GWAS for thyroid nodules, across DTC, thyroid nodules, and normal thyroid in a genome wide scan.

Expression quantitative trait loci analysis

We assessed the effect of associated genotypes on mRNA expression using the public expression quantitative trait loci (eQTL) database (47). Imputation of gene expression from the discovery series was performed using the PrediXcan package (https://github.com/hakyimlab/PrediXcan/tree/master/Software)(54).

The thyroid eQTL database (GTEx V6p, 278 thyroid samples) was used as a reference for imputation.

Statistical analysis

Genome wide associations were performed using PLINK version 1.90 beta (https://www.cog-genomics.org/plink/1.9), R statistics 3.2.3, and STATA software version 13.0 (StataCorp, College Station, Texas, USA). The genomic inflation factor (λ) for the series of GWAS was calculated to check the presence of population sub-structures. Quantile–quantile (*Q–Q*) plots by the qqman R package (https://cran.r-project.org/web/packages/qqman) were used to determine whether population stratification was adequately controlled. Manhattan plots were generated using Integrative Genomics Viewer (IGV) (http://software.broadinstitute.org/software/igv). In the series of GWAS, replications, and joint analysis, associations of each SNP were assessed by a logistic regression model adjusted with age and sex to estimate the per-allele odds ratios (ORs) with 95% confidence intervals (CIs) and *P*-values. We further performed a meta-analysis through the results of regression analysis from each of the three case-control groups using PLINK. Fixed-effects model was used to generate ORs with 95% CIs and the Cochran's Q and Higgins's I^2 tests were

applied to assess heterogeneity in meta-analysis (55).

Ethics statement

Informed consent was obtained from the Ansung cohort subjects for the discovery GWAS. Written informed consent was also obtained from participants of the replication study. This research protocol was approved by the Institutional Review Board of the National Cancer Center (IRB No. NCC2015-0238) and the Institutional Review Board of the Seoul National University Hospital (IRB No. H-1102-012-349 and H-1108-041-372).

Results

Discovery GWAS

In the discovery series, the genomic inflation factor (λ) was 1.0, indicating that there was no inflation of *P*-values as a result of population stratification. *Q*–*Q* plots showed no deviations of the observed distribution from the expected null distribution (Figure 12). The genome-wide association result for thyroid nodules was demonstrated with Manhattan plots (Figure 13). In the discovery genome-wide scan, the most robust signal (rs9952940) located an intron of *EPB41L3* on chromosome 18 ($P = 1.83 \times 10^{-7}$) (Figure 13 and 14a). The second significant signal (rs3780873) was observed in an intron of *ITGB1* ($P = 5.54 \times 10^{-6}$). In total, 29 candidate SNPs in 27 independent loci were selected for the validation study (Table 16).

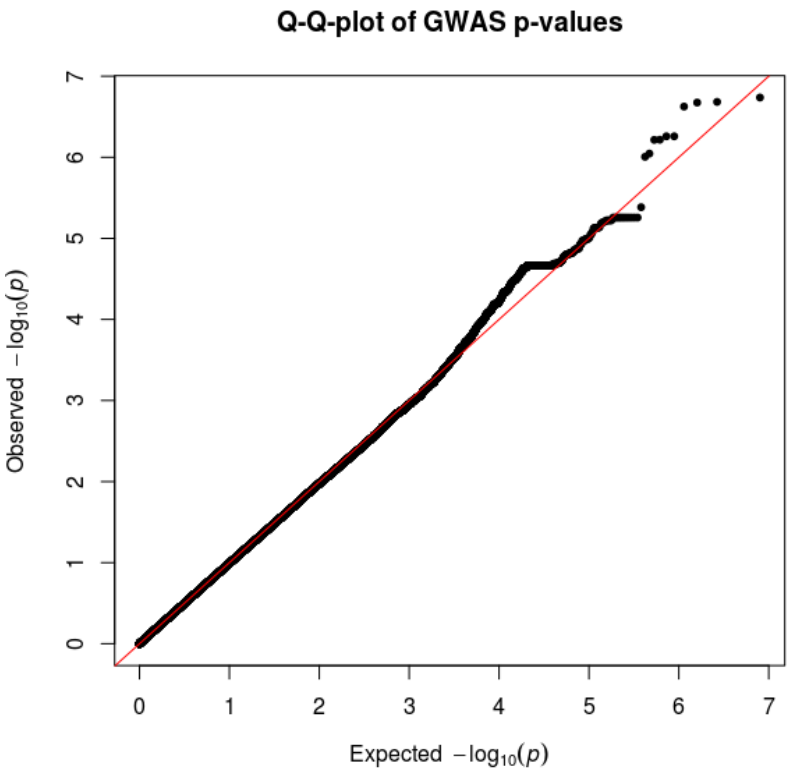

Figure 12. Q-Q plot of thyroid nodule GWAS result

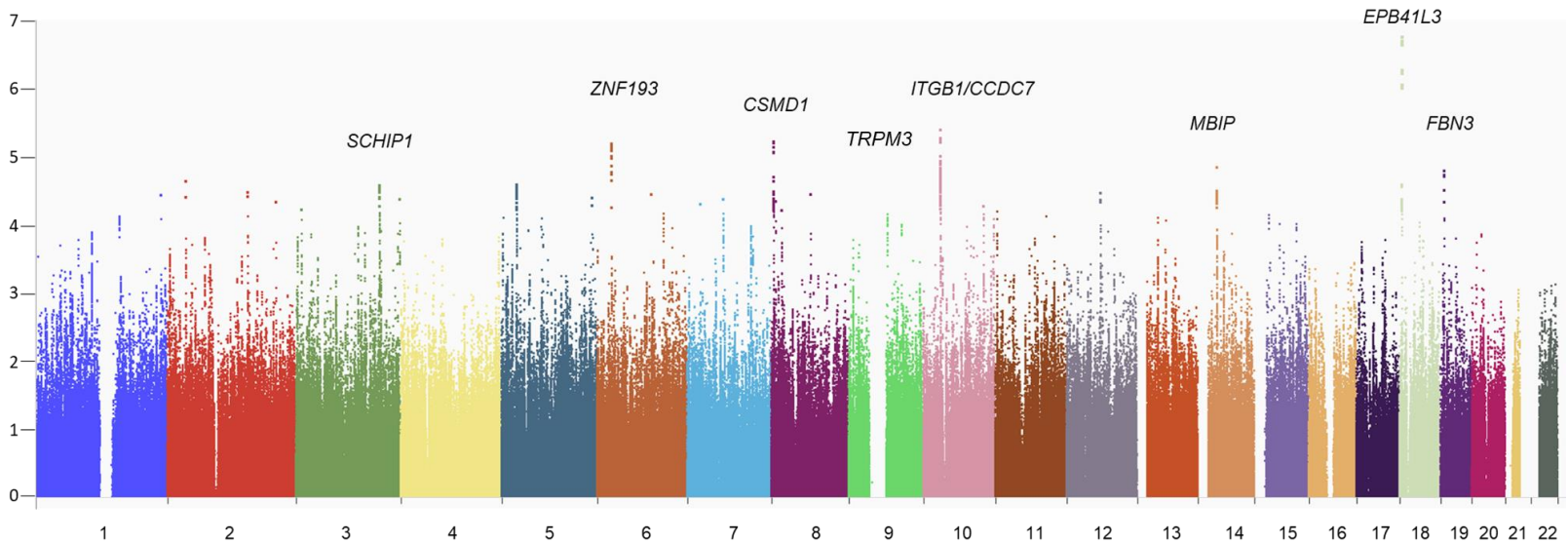

Figure 13. Manhattan plot for thyroid nodule GWAS

(a) (b) (c)

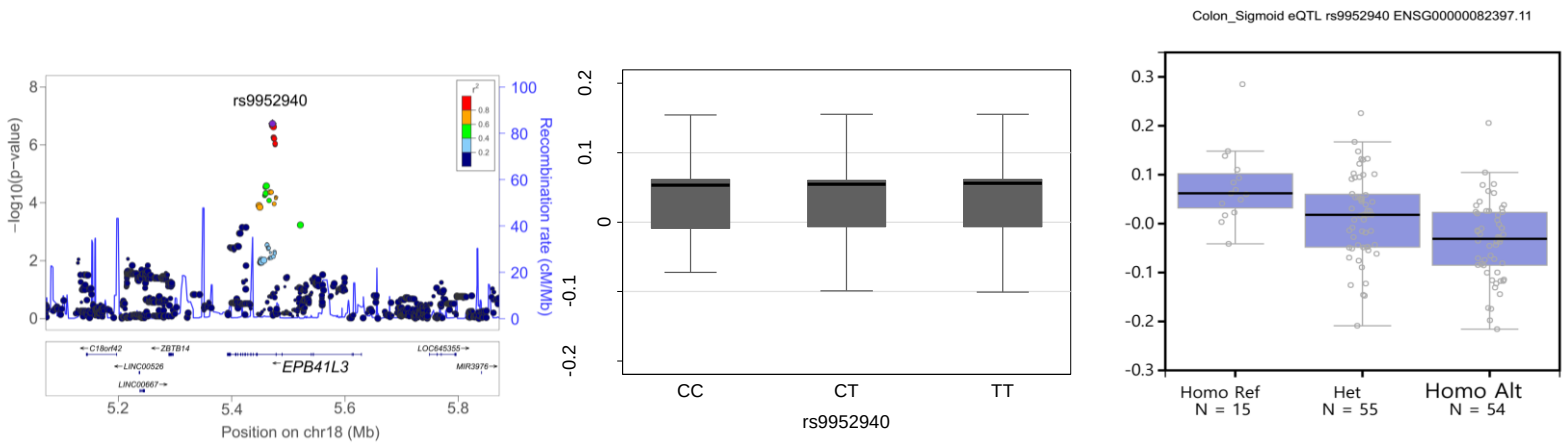

Figure 14. Regional association plots and expression of rs9952940 in *EPB41L3*

(a) Regional association plot in *EPB41L3*

(b) Imputed expression of *EPB41L3* (ENSG00000082397.11) according to rs9952940 genotype

(c) *EPB41L3* expression of colon tissue according to rs9952940 genotypes in GTEx data

Replication studies, joint analysis and meta-analysis

Among the 29 candidate SNPs, one signal, rs9952940 at *EPB41L3*, which showed the most robust association in the stage 1 genome scan, reached statistical significance in the first replication study (P = 0.045). The OR for rs4745021 at *TRPM3*, and rs2415317 and rs944289 at *MBIP/NKX2-1* in the replication study showed similar trends with the discovery GWAS, showing marginal significance (P-values 0.065, 0.087, and 0.095 respectively). We selected these SNPs for the second replication. In second replication study, variants at *TRPM3* and *MBIP/NKX2-1* were significantly associated with thyroid nodule (rs4745021; $P = 1.67 \times 10^{-4}$, rs2415317; $P = 0.043$, rs944289; $P = 0.028$). However, rs9952940 at *EPB41L3* was not replicated (P = 0.114). In joint analysis, rs4745021 at *TRPM3* ($P = 6.12 \times 10^{-8}$) did not reach the genome-wide significance threshold of $P < 5 \times 10^{-8}$ (Table 17 and Figure 15a). However, meta-analysis showed statistical significance (OR = 1.28, (95% CI; 1.18-1.40), $P = 2.11 \times 10^{-8}$, $Q = 0.41$, $I^2 = 0\%$). Three other SNPs did not reach genome-wide significance ($P < 5 \times 10^{-8}$) in the joint analysis (rs2415317; $P = 4.62 \times 10^{-5}$, rs944289; $P = 8.68 \times 10^{-5}$, rs9952940; $P = 1.42 \times 10^{-5}$) or meta-analysis (rs2415317; $P = 1.09 \times 10^{-5}$, rs944289; $P = 1.52 \times 10^{-5}$, rs9952940; $P = 5.24 \times 10^{-6}$) (Table 17). Considering the relatively small sample size, associations at *MBIP/NKX2-1*, which showed a marginal and a nominal significance in the first

and second replication, respectively, might be suggestive loci for thyroid nodules. The MAF of the four SNPs (rs4745021, rs2415317, rs944289 and rs9952940) were similar to those of the Asian population in the 1,000 genomes database (Table 18). The signal on *MSRB3,* which is a newly identified susceptibility locus for DTC from our previous GWAS, was not replicated in the first validation (P = 0.479) (Table 16).

Table 17. Suggestive associations from the GWAS for thyroid nodules

Chr SNP Position	Nearest Genes Minor/major allele	Series	Minor allele frequency Case/Control	OR (95% CI)	Q	I^2 (%)	P-value
9q21.13	Intron *TRPM3*	Discovery	0.27/0.21	1.42(1.19-1.70)			7.90×10^{-5}
rs4745021	T/A	First replication	0.26/0.22	1.26(0.99-1.60)			0.065
73267607		Second replication	0.28/0.24	1.24(1.11-1.38)			1.67×10^{-4}
		Joint	0.27/0.23	1.26(1.16-1.37)			6.12×10^{-8}
		Meta-analysis		1.28(1.18-1.40)	0.41	0	$*2.11\times10^{-8}$
14q13.3	158kb 3' of *MBIP*	Discovery	0.49/0.42	1.40(1.20-1.64)			1.45×10^{-5}
rs2415317	A/G	First replication	0.45/0.42	1.21(0.97-1.50)			0.087
36609678		Second replication	0.45/0.43	1.11(1.00-1.22)			0.043
		Joint	0.46/0.43	1.17(1.08-1.26)			4.62×10^{-5}
		Meta-analysis		1.19(1.10-1.29)	0.04	69	$*1.09\times10^{-5}$
14q13.3	119kb 3' of *MBIP*	Discovery	0.49/0.43	1.37(1.18-1.60)			5.72×10^{-5}
rs944289	T/C	First replication	0.46/0.43	1.20(0.97-1.48)			0.095
36649246		Second replication	0.46/0.44	1.11(1.01-1.23)			0.028
		Joint	0.47/0.44	1.16(1.08-1.24)			8.68×10^{-5}
		Meta-analysis		1.18(1.10-1.27)	0.08	61	$*1.52\times10^{-5}$
18p11.31	Intron *EPB41L3*	Discovery	0.19/0.27	0.61(0.50-0.73)			1.83×10^{-7}
rs9952940	C/T	First replication	0.21/0.24	0.77(0.60-0.99)			0.045
5472021		Second replication	0.20/0.23	0.91(0.81-1.02)			0.114
		Joint	0.20/0.24	0.82(0.75-0.90)			1.42×10^{-5}
		Meta-analysis		0.81(0.73-0.88)	<0.01	85	$*5.24\times10^{-6}$

Chr, Chromosome; CI, confidence interval; OR, odds ratio; SNP. Single-nucleotide polymorphism
The SNP positions are indexed to the National Center for Biotechnology Information (NCBI) build 37.

*P-value of meta-analysis with fixed-effects model

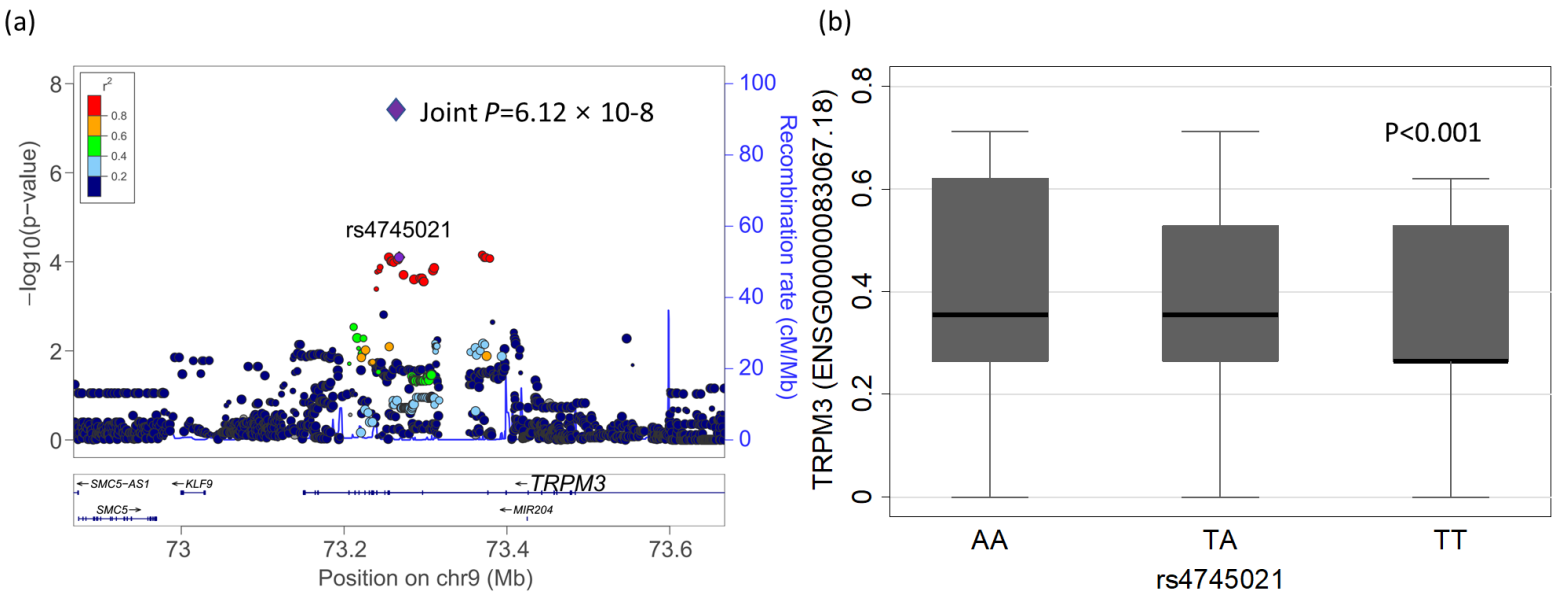

Figure 15. (a) Regional association plot for *TRPM3* (b) Predicted expression of *TRPM3* according to the rs4745021 genotype

Table 18. Comparison of the minor allele frequency between population of 1000 Genome and this study.

Chr	SNP	Gene	Alternative/ Reference allele	Risk allele frequency in 1000 Genome				Risk allele frequency in this study		
				African	American	European	Asian	Cases	Controls	Allelic OR
9	rs4745021	*TRPM3*	T/A	0.13	0.13	0.10	0.24	0.27	0.22	1.35
14	rs2415317	*MBIP/NKX2-1*	A/G	0.17	0.44	0.59	0.45	0.47	0.42	1.30
14	rs944289	*MBIP/NKX2-1*	T/C	0.15	0.44	0.59	0.45	0.48	0.43	1.27
18	rs9952940	*EPB41L3*	C/T	0.35	0.25	0.34	0.21	0.20	0.26	0.69

Chr, Chromosome; OR, odds ratio; SNP. Single-nucleotide polymorphism

Comparison of allele frequencies between DTC, thyroid nodules, and normal thyroid

We performed an additional analysis comparing the allele frequencies of 15 DTC susceptibility SNPs discovered in GWAS for DTC, and two candidate SNPs (rs4745021 at *TRPM3*, and rs2415317 at *MBIP/NKX2-1*) of thyroid nodules from this study among subjects with DTC, thyroid nodules, or normal thyroid (Table 19). Among them, two thyroid nodule-associated SNPs (rs4745021 at *TRPM3*, and rs2415317 at *MBIP/NKX2-1*), two thyroid cancer-associated SNPs (rs11175834 in *MSRB3*, and rs34081847 at *MBIP/NKX2-1*), and a SNP associated with both cancer and nodule (rs944289 at *MBIP/NKX2-1*) showed similar allele frequencies between the DTC and thyroid nodules. These results suggest that these signals might be shared susceptible loci between DTC and thyroid nodules, suggesting their common genetic etiology. Among these, signals in *MBIP/NKX2-1* (Figure 16 a-c) are considered to be common susceptibility loci for thyroid cancer and thyroid nodules because their associations with DTC risk have been repeatedly validated in GWAS for DTC, and the association with risk of thyroid nodules was replicated in this 3-stage GWAS. A signal in *MSRB3* (rs11175834) showed significant association with both thyroid cancer and thyroid nodules in the genome wide scan. However, the SNP failed to show any association in the replication series of thyroid nodules.

Signal in *TRPM3* were not candidate SNP in study on DTC. The other susceptible loci for thyroid cancers such as *NRG1* (re 6996585, Figure 16 d-f) and *FOXE1* showed no association with thyroid nodules and normal thyroid (Table 19), suggesting genetic susceptibilities for thyroid nodules to be distinct from thyroid cancer.

Table 19. Subgroup analysis comparing allele frequencies of DTC or thyroid nodule susceptibility SNPs between DTC, thyroid nodules, and normal thyroid in a genome wide scan

					Minor allele frequency			DTC vs. Normal thyroid		DTC vs. Thyroid nodule		Thyroid nodule vs. Normal thyroid	
Chr	SNP	Gene	Position	Minor/Major allele	DTC (n=470)	Thyroid nodule (n=811)	Normal Thyroid (n=691)	OR	P	OR	P	OR	P
DTC susceptibility SNPs													
1	rs4915076	*VAV3*	108359505	C/T	0.23	0.31	0.32	0.57	2.29×10^{-6}	0.60	9.00×10^{-6}	0.95	0.58
1	rs4649295	*PCNXL2*	233416538	T/C	0.12	0.19	0.17	0.64	1.91×10^{-3}	0.54	1.47×10^{-5}	1.12	0.28
2	rs1549738	*DIRC3*	218118722	G/A	0.39	0.45	0.47	0.70	4.62×10^{-4}	0.75	3.67×10^{-3}	0.91	0.23
2	rs12990503	*DIRC3*	218294217	G/C	0.31	0.36	0.38	0.80	3.70×10^{-2}	0.86	0.14	0.92	0.30
3	rs9858271	*FHIT*	59545330	G/A	0.50	0.43	0.43	1.31	7.20×10^{-3}	1.46	1.56×10^{-4}	1.01	0.92
4	rs1874564	*SEPT11*	77858105	A/G	0.23	0.31	0.32	0.61	1.35×10^{-5}	0.63	2.57×10^{-5}	0.94	0.48
8	rs12542743	*NRG1*	32318355	C/T	0.32	0.26	0.25	1.50	3.22×10^{-4}	1.42	1.17×10^{-3}	1.08	0.38
8	rs6996585	*NRG1*	32400803	G/A	0.30	0.23	0.21	1.59	6.38×10^{-5}	1.45	6.87×10^{-4}	1.15	0.15
8	rs2439302	*NRG1*	32432369	G/C	0.27	0.22	0.20	1.41	4.33×10^{-3}	1.27	3.59×10^{-2}	1.14	0.17
9	rs72753537	*FOXE1*	100660746	C/T	0.12	0.08	0.07	1.61	4.25×10^{-3}	1.57	3.13×10^{-3}	1.05	0.75
12	rs11175834	*MSRB3*	65992636	T/C	0.21	0.18	0.12	1.80	1.82×10^{-5}	1.11	0.39	1.58	4.69×10^{-5}
12	rs16934253	*SLC24A6*	113737225	A/G	0.05	0.02	0.02	2.37	2.14×10^{-3}	2.69	2.41×10^{-4}	0.84	0.52
14	rs34081947	*MBIP/NKX2-1*	36559531	T/C	0.47	0.46	0.40	1.51	6.69×10^{-5}	1.01	0.96	1.34	2.19×10^{-4}
14	rs944289	*MBIP/NKX2-1*	36649246	T/C	0.51	0.49	0.43	1.48	1.20×10^{-4}	0.96	0.69	1.37	5.72×10^{-5}
19	rs7248104	*INSR*	7224431	A/G	0.43	0.37	0.32	1.66	1.55×10^{-6}	1.22	0.04	1.30	1.14×10^{-3}
Thyroid nodule susceptibility SNPs													
9	rs4745021	*TRPM3*	73267607	T/A	0.26	0.27	0.21	1.35	0.01	0.95	0.62	1.42	7.90×10^{-5}
14	rs2415317	*MIBP/NKX2-1*	36609678	A/G	0.51	0.49	0.42	1.56	1.63×10^{-5}	0.98	0.81	1.40	1.45×10^{-5}

Chr, Chromosome; DTC, differentiated thyroid cancer; OR, odds ratio; SNP. Single-nucleotide polymorphism
The SNP positions are indexed to the National Center for Biotechnology Information (NCBI) build 37.

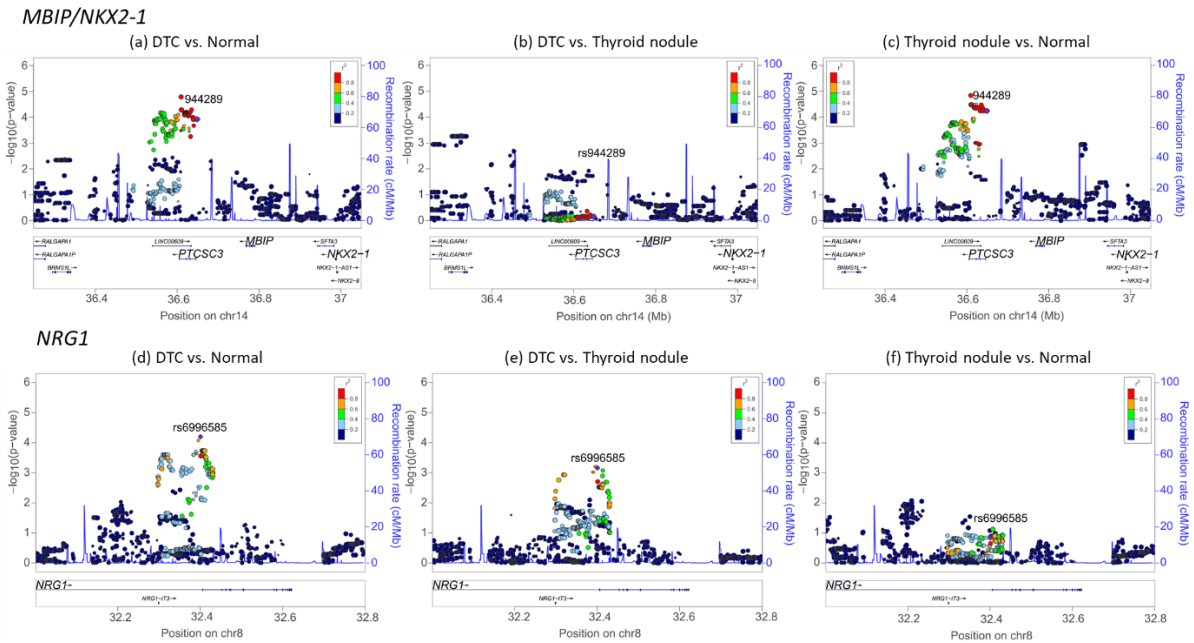

Figure 16. Regional association plots for *MBIP/NKX2-1* (DTC and thyroid nodule-risk locus) and *NRG1* (known DTC risk locus). Association plots for the *MBIP/NKX2-1* region comparing between (a) DTC and normal thyroid, (b) DTC and thyroid nodules, and (c) thyroid nodules and normal thyroid, respectively. Association plots for the *NRG1* region comparing between (d) DTC and normal thyroid, (e) DTC and thyroid nodules, and (f) thyroid nodules and normal thyroid, respectively.

Expression quantitative trait loci analysis

Imputed expression of the discovery series using PrediXcan revealed that the polymorphism rs4745021 have an eQTL for *TRPM3* (ENSG00000083067.18) in thyroid tissue (Figure 15b, and Table 20). Expression data for *EPB41L3* (ENSG00000082397.11) did not show a significant difference among the rs9952940 genotypes at *EPB41L3* (Figure 14b, and Table 20). Although *EPB41L3* expression was significantly different according to the rs9952940 genotype in colon tissues (Figure 14c). The expression data for *MBIP*, *NKX2-1*, *BRMS1L*, *PTCSC3*, *RP11-116N8.4*, or *LINC00609* that are located near *MIBP/NKX2-1*, by rs2415317 and rs944289 genotypes were not available from GTEx.

Table 20. eQTL results of candidate SNPs and nearby genes

Chr	SNP	Gene/Ensembl Gene Record	Genotype			P-value
9	rs4745021	*TRPM3*	AA(n=874)	AT(n=524)	TT(n=104)	
		ENSG00000083067.18	0.400±0.204	0.371±0.191	0.299±0.188	<0.001
18	rs9952940	*EPB41L3*	CC(n=65)	CT(n=548)	TT(n=889)	
		ENSG00000082397.11	0.030±0.065	0.033±0.059	0.037±0.061	0.358

Chr, Chromosome; SNP. Single-nucleotide polymorphism

Discussion

GWAS for DTC

In this two stage GWAS for DTC, we confirmed the associations of signals at *NRG1, NKX2-1, FOXE1* and *DIRC3.* Moreover, novel DTC susceptibility loci on *VAV3, PCNXL2, FHIT, SEPT11, MSRB3* and *INSR* were identified.

A variant (rs2439302) on *NRG1* was reported to be associated with thyroid stimulating hormone (TSH) level and DTC in previous GWASs (30, 56). Subsequently, several replication studies validated the association between rs2439302 and thyroid cancer (Table 1) (40, 41, 57). Neuregulin 1, which is encoded by the *NRG1* gene and acts on the *ERBB* family of tyrosine kinase receptors, could behave as a tumor suppressor in breast cancer cells (58). One study showed that neuregulin 1 promotes the proliferation and self-renewal of *HER2*-low breast cancers (59). The intrinsic resistance of PTC to a *BRAF* inhibitor is accompanied by increased *ERBB3* (*HER3*) signaling, which is dependent on *NRG1* autocrine signaling (60). This *NRG1* dysregulation is closely linked to *PI3K-AKT* and *MAPK* signaling pathway via *ERBB* in lung cancer (61). Thus, the upregulated *NRG1* expression could be associated with thyroid cancer development, especially in $BRAF^{V600E}$ mutation positive PTC.

In our GWAS, the most robust association was located near *NRG1.* Therefore,

we could suppose the *NRG1* loci might be the susceptible loci for PTC in this population. For instance, Gudmundsson et al. reported that a significant correlation was observed between genotypes on *NRG1* and relative expression of *NRG1* in blood (30), and we hypothesized that the expression level of *NRG1* in thyroid tissue might be determined by the identified variants. As expected, we also found that *cis*-eQTL data showed that those variants in the *NRG1* region were associated with expression of *NRG1* in both normal and cancer thyroid tissues from our RNA sequencing results. It was the same when we analyzed using the predicted expression results or public database, and those variants were associated with an increased *NRG1* expression in normal thyroid and whole blood. Furthermore, the GSEA identified that one of the variants rs6996585 was associated with many pathways related to cellular growth or cancer, and the ERBB-MAPK signalling pathway was the most significantly enriched with its related signals (Figure 7a and Table 10). Our clinical results also showed that the variants of *NRG1* are associated with lymph node metastasis of thyroid cancer, especially in $BRAF^{V600E}$ mutated PTC. A recent report supported this association showing an association between a variant of *NRG1* (rs2439302) and lymph node metastasis in PTC (38). Although we did not demonstrate the direct effects of the increased expression of *NRG1* on tumour aggressiveness, we could postulate a possibility that the increased

expression of *NRG1* in the thyroid tissue, which is associated with identified variants, might influence the development or progression of thyroid cancer. That is, the *NRG1* region could be an important risk gene for the susceptibility or prognosis of thyroid cancer. While it is uncertain why the *NRG1* region demonstrated the most significant association in our study, one plausible explanation could be that it was caused by the difference in MAF of the *NRG1* variant between the Asian population and other populations (Table 4).

A Mutation in *NKX2-1*, known as thyroid transcription factor (*TTF) 1*, is a causative mutation for brain-lung-thyroid syndrome characterized by congenital hypothyroidism, respiratory distress syndrome, and benign hereditary chorea (62). In previous GWAS for DTC, one common variant (rs944289), located near *MBIP/NKX2-1*, was associated with the mentioned diseases (29, 30) and several studies validated the association (40, 42). Jendrzejewski et al. reported that expression of noncoding RNA gene named papillary thyroid carcinoma susceptibility candidate 3 (*PTCSC3*) located near *MBIP/NKX2-1* was down-regulated in thyroid tumour tissue, and the risk allele (T) was associated with the profound suppression, which implies that *PTCSC3* could have some role as a tumor suppressor in DTC (63). We confirmed that the variant in rs944289 and another variant in rs34081947 were related with PTC, and the expression of *NKX2-1* was increased in thyroid tissues containing

those variants (Table 6).

FOXE1 (Forkhead Box E1), also known as *TTF2,* plays a role in the development or differentiation of thyroid (64, 65). Landa et al. demonstrated that variants near *FOXE1* affected *FOXE1* transcription through recruitment of *USF1/USF2* transcription factors (37). The *FOXE1* locus was reported as a major genetic determinant of risk in DTC and radiation-related PTC in several GWASs of European descents. Specifically, the relationship between common variations of *FOXE1* and DTC was validated in several studies (29, 36, 66), and rs965513 of *FOXE1* is the most well-replicated susceptibility locus in East-Asian population (Table 1). In recent research, some variants on *FOXE1* locus were associated with clinical phenotypes of PTC, such as tumour stage, size, lymphocytic infiltration and extrathyroidal extension (38, 39). Our results showed that only one variant (rs72753537) of *FOXE1* was significantly associated with PTC after replication, but there exist possibilities of 6 other variants as risk loci of DTC development (Table 12). We also could not find any association between the variants and the *FOXE1* expression levels or clinical phenotypes despite the fact that the variants' association with *FOXE1* expression level was observed in the public database. *FOXE1* is known as the most susceptible gene of DTC. However, the effect of *FOXE1* variants on thyroid cancer risk seemed to be less significant than that of *NRG1* in this Asian

population. Compared with 1000genome database, there are significant ethnic differences in allele frequencies for variants of *FOXE1* between the European and Asian populations (MAF; 0.14-0.34 vs. 0.08-0.13). This suggests that Asian populations with this polymorphism are relatively small and are thus less susceptible to thyroid cancer than European populations.

Given that the association between rs966423 on the *DIRC3* region and thyroid cancer was first reported in a European GWAS (30) and were replicated in limited ethnic groups, it is impossible to ignore the possibility of the presence of population heterogeneity (32). Wang et al. showed that the variant rs966423 on *DIRC3* was associated with increased PTC risk (40), which was confirmed by our present study ($P = 0.0067$, Supplementary Table 13). In a recent report, rs966423 was associated with increased mortality in DTC (67). In this study, two SNPs were associated with DTC that were different SNPs from the previous report. Notably, rs1549438 was associated with both the FTC and the expression of *TNS1*.

As described above, the previous reported region of *NKX2-1, FOXE1* and *DIRC3* showed either a relatively less association than a European study, or the association was observed in different SNPs. However, these regions were still found to be risk loci in Korean or other Asian populations as in European populations. As for the regions in *IMMP2L, DHX35, ARSB* and *WDR11-AS1,*

no association was observed (Figure 10 and Table 13).

Here, we identified six novel susceptible loci near *VAV3, PCNXL2, INSR, MRSB3, FHIT* and *SEPT11* that were no found in previous European studies.

VAV3 gene encodes a guanine nucleotide exchange factor 3 for Rho family GTPases, which activates pathways involving in actin cytoskeletal rearrangements and transcriptional alterations (68). Casual variants on *VAV3* were associated with hypothyroidism in both a European and our previous Korean GWAS (69, 70). *VAV3* expression in thyroid cancer cell is also known to be RET/PTC and *RAS* mutation-specific because *VAV3* is involved in *PI3K* signaling and subsequent *AKT* activation (71). In this study, we first demonstrated that the association between the variation at *VAV3* and DTC risk or *VAV3* expression.

MSRB3 encodes a zinc-containing methionine sulfoxide reductase B3 and its mutations are associated with human deafness (72). Methionine sulfoxide reductase is suggested to utilize catalytic selenocysteine (73). Selenocysteine is an essential component of deiodinases enzymes, which convert thyroxine(T4) into triiodothyronine(T3), and is also associated with thyroid autoimmunity (74). Thus, variants on *MSRB3* could be related with pathogenesis of thyroid cancer. The fusion of *MSRB3* was found in the primary and metastasis tumour of spindle cell (75). Although we could not find any association between the

variant of *MSRB3* and *MSRB3* expression or clinical phenotypes, there was an association with FTC as well as with PTC.

Another newly identified susceptibility locus was located near *INSR*. In a previous meta-analysis of GWASs for thyroid-related traits, common variants in *INSR* were reported to be associated with the levels of TSH (76). Insulin is reported to be upregulated in various cancers (77, 78), and in *in vitro* studies, the expression of *INSR* was also elevated in malignant thyrocytes (79). In this study, we found a risk variant of PTC in *INSR* gene, although it was not so strong. However, the presence of a risk allele was significantly associated with lymph node metastasis, suggesting a possible pathogenic role in thyroid cancer.

The other 3 new genes are *FHIT* (Fragile Histidine Triad), *PCNXL2,* and *SEPT11*. *FHIT* encodes a diadenosine 5',5'''-P1,P3-triphosphate hydrolase, and is known as a tumour suppressor gene in a variety of common human cancers (80, 81). In thyroid cancer, the inactivation of *FHIT* was suggested to be associated with the pathogenesis of thyroid neoplasm (82), and the homozygous deletion and promoter methylation of *FHIT* was reported to be associated with DTC (83, 84). This is concordant with our result, although the association between a variant in *FHIT* and PTC risk was not so strong, and no association with the gene expression or clinical phenotypes was observed.

SEPT11 encodes septin-11, which is involved in filament-forming cytoskeletal

GTPase, and may play a potential role in cytokinesis (85). *SEPT11* is also known as a gene of the fusion partners of MLL in chronic neutrophilic leukemia (86). Regarding *PCNXL2,* there was one report suggesting that it might have a role in tumorigenesis of colorectal carcinomas with a high microsatellite instability (87). Biological functions of both *PCNXL2* and *SEPT11* on thyroid tissue were not defined. However, the expression of *PCNXL2* was increased in thyroid tissues with a candidate variant, and there was an association between a candidate variant of *SEPT11* and extrathyroidal extension, suggesting their possible role in the development of thyroid cancer.

Recently, Gudmundsson et al. reported that a meta-analysis of GWAS yields five novel risk loci for thyroid cancer(35). Among the five loci, one locus (1q42.2, *PCNXL2*) was included in the significant loci in our result (Table 5). However, the intron variant (rs12129938) of *PCNXL2*, which was reported as most significant by Gudmundsson et al., showed only marginal association (P = 0.002) in our result. Instead, the other intronic variant (rs4649295) showed most significant association (Joint P = 6.0×10^{-8}). In addition, 10q24.33 locus (near *OBFC1*) showed suggestive association (P = $8.72 \times 10\text{-}6$) in our discovery result. The other novel risk loci (3q26.2, 5q22.1 and 15q22.33) and one replicating locus (5p15.33) showed no associated signal (P < 0.001) in our result. The comparison of reported SNPs with our result are summarized in Table 14.

The limitation of this study is that we used two different genotyping platforms in the discovery stage because the genotyping of the control sample was performed in the past. As a consequence, potential SNPs were excluded from the study to remove batch effect, and information might have been lost in the process. Another limitation was the imbalance of gender between the case and control groups caused by a higher prevalence of thyroid cancer in the female population. To adjust for this imbalance, we repeated the test with stratification for sex, and although the statistical significance was lowered, there was no difference in candidate genes. Lastly, the control samples of replication stage were from participants of a relatively higher age. However, it was an inevitable consequence of ensuring that there is no misclassification of participants (to confirm that the control participants truly did not have cancer). Furthermore, these participants went through ultrasonography examination to ensure they are "super normal" controls, and thus the result of comparison between cancer and normal should thus be more reliable and be a unique value of this study.

The strength of this study is that this is the first study to be performed in Asian population using GWAS in thyroid cancer, and we identified some ethnic differences. When we compared the DTC SNPs reported in Europeans with our Stage 1 result, we found a similar effect size in SNPs of *FOXE1* and *NRG1* between the two populations, but the effect of *NKX2-1* and *DIRC3* were lower

in Koreans compared to Europeans, and we suspect this may be due to a difference of genetic susceptibility between different ethnic populations.

In addition, we discovered specific variants of DTC that have different effects on PTC from their effects on FTC. Firstly, the variants of *NRG1* were more associated with PTC than DTC, showing the most robust effect. Secondly, variant of *SLC24A6* showing a high-risk effect (OR = 3.32) While both PTC and FTC are cancers of follicular cell origin, the mutational profiles are quite different between PTC and FTC, our results suggest that the risk assessment of thyroid cancer development should be tailored according to cancer type and personal genotypes. Furthermore, we were the first to identify that the candidate SNPs influence the molecular and biological changes on the development of thyroid cancer by performing eQTL analysis.

GWAS for Thyroid nodule

In this 3-stage GWAS, the signal in *TRPM3* showed the most robust association with thyroid nodules and those in *MIBP/NKX2-1* also demonstrated a possible association. A variant at *TRPM3* reached the genome-wide significance threshold in meta-analysis. However other associations did not show genome wide significance in the joint analysis or meta-analysis because of the small number of subjects. Our findings provide evidence that thyroid nodules may arise from genetic predispositions. Candidate SNPs of thyroid nodule showed a similar trend of association with thyroid cancer; however, SNPs with the most risk of thyroid cancer including *NRG1* and *FOXE1* did not show any association with thyroid nodules, suggesting a distinct genetic susceptibility of thyroid nodules from thyroid cancer.

In epidemiologic studies, most thyroid diseases including thyroid cancer, Graves' disease, and hypothyroidism have been reported to have a large degree of genetic heritability. Approximately 50–70% of the predisposition to autoimmune thyroid disease was reported to arise from genetic factors (88), and inherited genetic factors for thyroid cancer development accounted for 53% (89). GWAS have been widely applied in the field of thyroid diseases for a decade to identify these genetic factors. Despite the accumulated data of genetic studies on thyroid disease, a GWAS for thyroid nodules has not been conducted

till date. I hypothesized that genetic factors could contribute to the development of thyroid nodules, which could share a common genetic predisposition with thyroid cancer. In this study, *TRPM3* was identified as susceptibility locus, and *MBIP/NKX2-1* as suggestive risk locus for thyroid nodules.

A variant on *TRPM3* (9q21.13) showed the most robust association in the joint analysis. *TRPM3* encodes Transient Receptor Potential Cation Channel Subfamily M member 3, which is a cation-selective channel related to cellular calcium signaling and homeostasis (90). *TRPM3* was also documented as component of an ionotropic steroid receptor in insulin-producing beta cells (91). The biological function of *TRPM3* in the thyroid gland is not well defined. Recently, *TRPM3* was reported to promote the growth of renal cell carcinoma (92). Thus, *TRPM3* expression could also affect tumorigenesis in thyroid cells.

Interestingly, in this discovery genome-wide scan, rs2415317 and rs944289 signals were detected near *MBIP/NKX2-1,* which is reported to be associated with DTC (29, 35, 93). Association of a following replication studies using an independent population showed a similar trend of OR with marginal or nominal significance. Wang et al. have also reported an association between benign thyroid nodules and rs944289 in a Chinese population (40). Therefore, although associations in *MBIP/NKX2-1* did not reach genome-wide significance, we could not completely exclude the possibility that variants on *MBIP/NKX2-1*

have associations with thyroid nodules. Thus, it could be suggested that variants on *MBIP/NKX2-1* might not be cancer-specific but play various roles in the neoplastic changes of follicular cells. In this analysis, rs2415317 and rs944289 also showed a similar association with thyroid cancer (Table 19). In addition to *MBIP/NKX2-1*, SNPs on *TRPM3*, *EPB41L3*, and *MSRB3* demonstrated similar allele frequencies in thyroid nodules and thyroid cancer. Further studies are needed to confirm whether these variants affect both thyroid cancer and thyroid nodule development.

A signal in *EPB41L3* showed the most robust association in discovery GWAS and showed statistical significance only in the first replication study. In joint analysis, the association with *EPB41L3* did not reach genome-wide statistical significance. *EPB41L3* (18p11.31), Erythrocyte Membrane Protein Band 4.1 Like 3, encodes protein 4.1B, which is related with actin binding and structural constituents of the cytoskeleton (94). Protein 4.1B is expressed in most mammalian tissues and is prominently localized at the plasma membrane in regions of cell-cell contact (95). Down-regulation of protein 4.1B is frequently observed in tumors and it is considered as a potential tumor suppressor (96). Polymorphisms in *EPB41L3* were documented to be associated with kidney function and sleep duration in other GWAS (97, 98). Although, a cis-eQTL of *EPB41L3* was not identified, we could not completely

exclude the possibility that variants of *EPB41L3* could influence molecular and biological changes in the thyroid gland via trans-eQTL.

In GWAS for thyroid cancer, several susceptibility loci of DTC, including *FOXE1* (9q22.33), *MBIP/NKX2-1* (14q13.3), *DIRC3* (2q35), and *NRG1* (8p12) were validated in diverse populations, and other candidate loci (*IMMP2L, RARRES1, SNAPC4/CARD9, BAT,* and *DHX35*) were identified and replicated in specific ethnicities (29-32, 36). I previously described seven novel susceptibility loci (*VAV3, PCNXL2, INSR, MRSB3, FHIT, SEPT11,* and *SLC24A6*) associated with DTC in Chapter I. GWAS for goiters was also conducted and revealed that genetic loci near *CAPZB, FGF7,* and *LOC440389* were associated with thyroid volume (99). Regarding the GWAS for thyroid function, many susceptibility loci for thyrotropin levels (*PDE10A, PDE8B, VEGFA, CAPZB, NRG1, MBIP, SYN2, IGFBP2, NR3C3*, *FGF7*, etc.) and free thyroxine levels (*FOXE1, B4GALT6 LHX3, DIO1*, *AADAT*, etc.) have been identified (56, 76, 100, 101). In these series of GWAS for various thyroid-related traits, the common variants in several genetic loci including *MBIP/NXK2-1*, *FOXE1*, *VAV3*, *INSR*, and *NRG1* showed associations with multiple thyroid-related phenotypes, suggesting that thyroid disorders could be closely linked with each other via genetic predispositions (Figure 17). However, more studies are required to clarify the shared genetic traits of *TRPM3* and

MBIP/NXK2-1 between thyroid nodules and thyroid cancer.

Importantly, no susceptible locus other than *TRPM3,* and *MBIP/NXK2-1*, was identified in thyroid nodules in this GWAS, and the statistical significance of associations was not so evident in this study. Moreover, no association with previously known susceptible loci for thyroid cancer like *NRG1* or *FOXE1* was observed (Figure 17). These data suggest that the heritability of thyroid nodules may not be so strong, and that their genetic susceptibility is distinct from thyroid cancer.

Clinically, it is important to properly diagnose a malignancy among thyroid nodules. The probability of malignancy varies according to the sonographic characteristics (102), and so, fine needle aspiration cytology is used for diagnosis according to the recommendations (18, 103, 104) However, it is still inconclusive whether thyroid nodules should be followed-up to prevent missing an occurrence of thyroid cancer from nodules with benign cytologic results. If patients with benign thyroid nodules share a common genetic etiology with patients with thyroid cancer, thyroid nodules should be followed-up closely. In a long term follow-up study on thyroid nodules, the chance for thyroid cancer development from cytologically benign thyroid nodules was reported to be extremely low (105) and was considered to be similar to that of the general population (106). Thus, repeated evaluation for detecting thyroid cancer is not

recommended for benign thyroid nodules (107). Consistent with the clinical observations, the genetic traits of thyroid nodules found in this study seemed to differ from those of thyroid cancer. However, variants on *MBIP/NKX2-1* or a few other SNPs are possibly associated with both thyroid nodules and thyroid cancer. Further study is therefore needed to identify the common genetic factors of thyroid cancer and nodules, which could elucidate the association between thyroid cancer and nodule from a genetic perspective.

While the present study is the first GWAS for thyroid nodules, our study had limitations in that the number of subjects was small and a detailed pathology of thyroid nodules was not indicated. Thyroid nodules are not a single disease entity. Most benign thyroid nodules are pathologically diagnosed as hyperplastic nodules and follicular adenomas. Genetic predispositions between hyperplastic nodules and follicular adenomas could be different. Genetic factors could not be detected in this GWAS possibly because hyperplastic nodules and follicular adenomas might not share common genetic factors. However, it is not clinically feasible to distinguish thyroid pathology from features of ultrasonography or cytologic results of fine needle aspiration. In addition, the possibility of thyroid cancer being included among thyroid nodules cannot be totally excluded in these cases. Another limitation was that stratification analysis for age and sex could not performed due to the limitation

of the number of subjects.

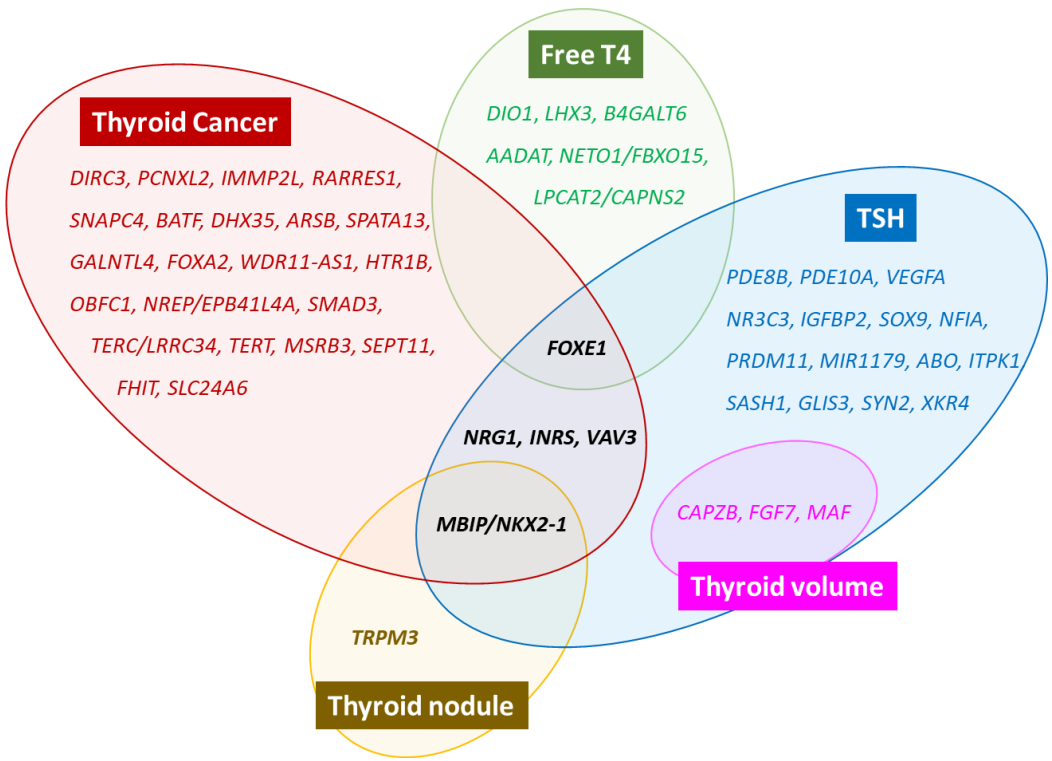

Figure 17. Genetic loci from the GWAS of thyroid-related traits

Summary and conclusions

Genome-wide association study for DTC was conducted in Koreans. Significant associations at previously reported loci of *NRG1, FOXE1, NKX2-1,* and *DIRC3* were confirmed, and *NRG1* was the most significantly associated in this Asian population. Novel susceptibility loci at *VAV3, PCNXL2, FHIT, SEPT11, MSRB3* and *INSR* were identified. These results were validated with *cis*-eQTL analysis using RNA sequencing data from the tumour and normal thyroid tissues.

And the first GWAS for thyroid nodules was also conducted and signals in *TRPM3* was identified as susceptibility locus. *MBIP/NXK2-1* could be possible risk locus associated with thyroid nodules. The genetic susceptibility for thyroid nodules was distinct from that for thyroid cancer, although these risk loci, especially on *MBIP/NKX2-1,* seemed to be shared between them.

I propose that these results can be implied for diagnosis and treatment of thyroid cancer and nodule and provide more insight into genetic factors in the era of personalized medicine in cancer.

Reference

1. Pellegriti G, Frasca F, Regalbuto C, Squatrito S, Vigneri R. Worldwide increasing incidence of thyroid cancer: update on epidemiology and risk factors. J Cancer Epidemiol. 2013;2013:965212.

2. Jung KW, Won YJ, Kong HJ, Lee ES. Cancer Statistics in Korea: Incidence, Mortality, Survival, and Prevalence in 2016. Cancer Res Treat. 2019;51(2):417-30.

3. Davies L, Morris LG, Haymart M, Chen AY, Goldenberg D, Morris J, et al. American Association of Clinical Endocrinologists and American College of Endocrinology Disease State Clinical Review: The Increasing Incidence of Thyroid Cancer. Endocr Pract. 2015;21(6):686-96.

4. Williams D. Radiation carcinogenesis: lessons from Chernobyl. Oncogene. 2008;27 Suppl 2:S9-18.

5. Zhao ZG, Guo XG, Ba CX, Wang W, Yang YY, Wang J, et al. Overweight, obesity and thyroid cancer risk: a meta-analysis of cohort studies. J Int Med Res. 2012;40(6):2041-50.

6. Kitahara CM, Platz EA, Freeman LE, Hsing AW, Linet MS, Park Y, et al. Obesity and thyroid cancer risk among U.S. men and women: a pooled analysis of five prospective studies. Cancer Epidemiol Biomarkers Prev. 2011;20(3):464-72.

7. Schmid D, Ricci C, Behrens G, Leitzmann MF. Adiposity and risk of thyroid cancer: a systematic review and meta-analysis. Obes Rev. 2015;16(12):1042-54.

8. Park YJ, Ahn HY, Choi HS, Kim KW, Park DJ, Cho BY. The long-term

outcomes of the second generation of familial nonmedullary thyroid carcinoma are more aggressive than sporadic cases. Thyroid. 2012;22(4):356-62.

9. Capezzone M, Marchisotta S, Cantara S, Busonero G, Brilli L, Pazaitou-Panayiotou K, et al. Familial non-medullary thyroid carcinoma displays the features of clinical anticipation suggestive of a distinct biological entity. Endocr Relat Cancer. 2008;15(4):1075-81.

10. Ito Y, Kakudo K, Hirokawa M, Fukushima M, Yabuta T, Tomoda C, et al. Biological behavior and prognosis of familial papillary thyroid carcinoma. Surgery. 2009;145(1):100-5.

11. Loh KC. Familial nonmedullary thyroid carcinoma: a meta-review of case series. Thyroid. 1997;7(1):107-13.

12. Uchino S, Noguchi S, Kawamoto H, Yamashita H, Watanabe S, Yamashita H, et al. Familial nonmedullary thyroid carcinoma characterized by multifocality and a high recurrence rate in a large study population. World J Surg. 2002;26(8):897-902.

13. Sturgeon C, Clark OH. Familial nonmedullary thyroid cancer. Thyroid. 2005;15(6):588-93.

14. Nose V. Familial non-medullary thyroid carcinoma: an update. Endocr Pathol. 2008;19(4):226-40.

15. Park YJ, Ahn HY, Choi HS, Kim KW, Park do J, Cho BY. The long-term outcomes of the second generation of familial nonmedullary thyroid carcinoma are more aggressive than sporadic cases. Thyroid. 2012;22(4):356-62.

16. Peiling Yang S, Ngeow J. Familial non-medullary thyroid cancer: unraveling

the genetic maze. Endocr Relat Cancer. 2016;23(12):R577-R95.

17. Gara SK, Jia L, Merino MJ, Agarwal SK, Zhang L, Cam M, et al. Germline HABP2 Mutation Causing Familial Nonmedullary Thyroid Cancer. N Engl J Med. 2015;373(5):448-55.

18. Haugen BR. 2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer: What is new and what has changed? Cancer. 2016.

19. Knudsen N, Laurberg P, Perrild H, Bulow I, Ovesen L, Jorgensen T. Risk factors for goiter and thyroid nodules. Thyroid. 2002;12(10):879-88.

20. Kwong N, Medici M, Angell TE, Liu X, Marqusee E, Cibas ES, et al. The Influence of Patient Age on Thyroid Nodule Formation, Multinodularity, and Thyroid Cancer Risk. J Clin Endocrinol Metab. 2015;100(12):4434-40.

21. Jiang H, Tian Y, Yan W, Kong Y, Wang H, Wang A, et al. The Prevalence of Thyroid Nodules and an Analysis of Related Lifestyle Factors in Beijing Communities. Int J Environ Res Public Health. 2016;13(4):442.

22. Dean DS, Gharib H. Epidemiology of thyroid nodules. Best Pract Res Clin Endocrinol Metab. 2008;22(6):901-11.

23. Moon JH, Hyun MK, Lee JY, Shim JI, Kim TH, Choi HS, et al. Prevalence of thyroid nodules and their associated clinical parameters: a large-scale, multicenter-based health checkup study. Korean J Intern Med. 2017.

24. Brix TH, Kyvik KO, Hegedus L. Major role of genes in the etiology of simple goiter in females: a population-based twin study. J Clin Endocrinol Metab.

1999;84(9):3071-5.

25. Memon A, Berrington De Gonzalez A, Luqmani Y, Suresh A. Family history of benign thyroid disease and cancer and risk of thyroid cancer. Eur J Cancer. 2004;40(5):754-60.

26. Leux C, Truong T, Petit C, Baron-Dubourdieu D, Guenel P. Family history of malignant and benign thyroid diseases and risk of thyroid cancer: a population-based case-control study in New Caledonia. Cancer Causes Control. 2012;23(5):745-55.

27. Panicker V. Genetics of thyroid function and disease. Clin Biochem Rev. 2011;32(4):165-75.

28. Wellcome Trust Case Control C. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007;447(7145):661-78.

29. Gudmundsson J, Sulem P, Gudbjartsson DF, Jonasson JG, Sigurdsson A, Bergthorsson JT, et al. Common variants on 9q22.33 and 14q13.3 predispose to thyroid cancer in European populations. Nat Genet. 2009;41(4):460-4.

30. Gudmundsson J, Sulem P, Gudbjartsson DF, Jonasson JG, Masson G, He H, et al. Discovery of common variants associated with low TSH levels and thyroid cancer risk. Nat Genet. 2012;44(3):319-22.

31. Kohler A, Chen B, Gemignani F, Elisei R, Romei C, Figlioli G, et al. Genome-wide association study on differentiated thyroid cancer. J Clin Endocrinol Metab. 2013;98(10):E1674-81.

32. Figlioli G, Kohler A, Chen B, Elisei R, Romei C, Cipollini M, et al. Novel

genome-wide association study-based candidate loci for differentiated thyroid cancer risk. J Clin Endocrinol Metab. 2014;99(10):E2084-92.

33. Figlioli G, Chen B, Elisei R, Romei C, Campo C, Cipollini M, et al. Novel genetic variants in differentiated thyroid cancer and assessment of the cumulative risk. Sci Rep. 2015;5:8922.

34. Mancikova V, Cruz R, Inglada-Perez L, Fernandez-Rozadilla C, Landa I, Cameselle-Teijeiro J, et al. Thyroid cancer GWAS identifies 10q26.12 and 6q14.1 as novel susceptibility loci and reveals genetic heterogeneity among populations. Int J Cancer. 2015;137(8):1870-8.

35. Gudmundsson J, Thorleifsson G, Sigurdsson JK, Stefansdottir L, Jonasson JG, Gudjonsson SA, et al. A genome-wide association study yields five novel thyroid cancer risk loci. Nat Commun. 2017;8:14517.

36. Takahashi M, Saenko VA, Rogounovitch TI, Kawaguchi T, Drozd VM, Takigawa-Imamura H, et al. The FOXE1 locus is a major genetic determinant for radiation-related thyroid carcinoma in Chernobyl. Hum Mol Genet. 2010;19(12):2516-23.

37. Landa I, Ruiz-Llorente S, Montero-Conde C, Inglada-Perez L, Schiavi F, Leskela S, et al. The variant rs1867277 in FOXE1 gene confers thyroid cancer susceptibility through the recruitment of USF1/USF2 transcription factors. PLoS Genet. 2009;5(9):e1000637.

38. Jendrzejewski JP, Liyanarachchi S, Nagy R, Senter L, Wakely PE, Jr., Thomas A, et al. Papillary thyroid carcinoma: Association between germline DNA

variant markers and clinical parameters. Thyroid. 2016.

39. Penna-Martinez M, Epp F, Kahles H, Ramos-Lopez E, Hinsch N, Hansmann ML, et al. FOXE1 association with differentiated thyroid cancer and its progression. Thyroid. 2014;24(5):845-51.

40. Wang YL, Feng SH, Guo SC, Wei WJ, Li DS, Wang Y, et al. Confirmation of papillary thyroid cancer susceptibility loci identified by genome-wide association studies of chromosomes 14q13, 9q22, 2q35 and 8p12 in a Chinese population. J Med Genet. 2013;50(10):689-95.

41. Rogounovitch TI, Bychkov A, Takahashi M, Mitsutake N, Nakashima M, Nikitski AV, et al. The common genetic variant rs944289 on chromosome 14q13.3 associates with risk of both malignant and benign thyroid tumors in the Japanese population. Thyroid. 2015;25(3):333-40.

42. Matsuse M, Takahashi M, Mitsutake N, Nishihara E, Hirokawa M, Kawaguchi T, et al. The FOXE1 and NKX2-1 loci are associated with susceptibility to papillary thyroid carcinoma in the Japanese population. J Med Genet. 2011;48(9):645-8.

43. Cho YS, Go MJ, Kim YJ, Heo JY, Oh JH, Ban HJ, et al. A large-scale genome-wide association study of Asian populations uncovers genetic factors influencing eight quantitative traits. Nat Genet. 2009;41(5):527-34.

44. Kim Y, Han BG, Ko GESg. Cohort Profile: The Korean Genome and Epidemiology Study (KoGES) Consortium. Int J Epidemiol. 2016.

45. Tian C, Gregersen PK, Seldin MF. Accounting for ancestry: population

substructure and genome-wide association studies. Hum Mol Genet. 2008;17(R2):R143-50.

46. Yoo SK, Lee S, Kim SJ, Jee HG, Kim BA, Cho H, et al. Comprehensive Analysis of the Transcriptional and Mutational Landscape of Follicular and Papillary Thyroid Cancers. PLoS Genet. 2016;12(8):e1006239.

47. Consortium GT. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science. 2015;348(6235):648-60.

48. Westra HJ, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nature genetics. 2013;45(10):1238-43.

49. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81(3):559-75.

50. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102(43):15545-50.

51. Lee C, Choe EK, Choi JM, Hwang Y, Lee Y, Park B, et al. Health and Prevention Enhancement (H-PEACE): a retrospective, population-based cohort study conducted at the Seoul National University Hospital Gangnam Center, Korea. BMJ Open. 2018;8(4):e019327.

52. Chung GE, Lee Y, Yim JY, Choe EK, Kwak MS, Yang JI, et al. Genetic Polymorphisms of PNPLA3 and SAMM50 Are Associated with Nonalcoholic Fatty

Liver Disease in a Korean Population. Gut Liver. 2018;12(3):316-23.

53. Choe EK, Lee Y, Cho JY, Choi SH, Park B, Lee JE, et al. Search for genetic factor association with cancer-free prostate-specific antigen level elevation on the basis of a genome-wide association study in the Korean population. Eur J Cancer Prev. 2017.

54. Gamazon ER, Wheeler HE, Shah KP, Mozaffari SV, Aquino-Michaels K, Carroll RJ, et al. A gene-based association method for mapping traits using reference transcriptome data. Nat Genet. 2015;47(9):1091-8.

55. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ. 2003;327(7414):557-60.

56. Malinowski JR, Denny JC, Bielinski SJ, Basford MA, Bradford Y, Peissig PL, et al. Genetic variants associated with serum thyroid stimulating hormone (TSH) levels in European Americans and African Americans from the eMERGE Network. PLoS One. 2014;9(12):e111301.

57. Liyanarachchi S, Wojcicka A, Li W, Czetwertynska M, Stachlewska E, Nagy R, et al. Cumulative risk impact of five genetic variants associated with papillary thyroid carcinoma. Thyroid. 2013;23(12):1532-40.

58. Chua YL, Ito Y, Pole JC, Newman S, Chin SF, Stein RC, et al. The NRG1 gene is frequently silenced by methylation in breast cancers and is a strong candidate for the 8p tumour suppressor gene. Oncogene. 2009;28(46):4041-52.

59. Lee CY, Lin Y, Bratman SV, Feng W, Kuo AH, Scheeren FA, et al. Neuregulin autocrine signaling promotes self-renewal of breast tumor-initiating cells by triggering HER2/HER3 activation. Cancer Res. 2014;74(1):341-52.

60. Montero-Conde C, Ruiz-Llorente S, Dominguez JM, Knauf JA, Viale A, Sherman EJ, et al. Relief of feedback inhibition of HER3 transcription by RAF and MEK inhibitors attenuates their antitumor effects in BRAF-mutant thyroid carcinomas. Cancer Discov. 2013;3(5):520-33.

61. Fernandez-Cuesta L, Thomas RK. Molecular Pathways: Targeting NRG1 Fusions in Lung Cancer. Clin Cancer Res. 2015;21(9):1989-94.

62. Shetty VB, Kiraly-Borri C, Lamont P, Bikker H, Choong CS. NKX2-1 mutations in brain-lung-thyroid syndrome: a case series of four patients. J Pediatr Endocrinol Metab. 2014;27(3-4):373-8.

63. Jendrzejewski J, He H, Radomska HS, Li W, Tomsic J, Liyanarachchi S, et al. The polymorphism rs944289 predisposes to papillary thyroid carcinoma through a large intergenic noncoding RNA gene of tumor suppressor type. Proc Natl Acad Sci U S A. 2012;109(22):8646-51.

64. Zannini M, Avantaggiato V, Biffali E, Arnone MI, Sato K, Pischetola M, et al. TTF-2, a new forkhead protein, shows a temporal expression in the developing thyroid which is consistent with a role in controlling the onset of differentiation. The EMBO journal. 1997;16(11):3185-97.

65. Parlato R, Rosica A, Rodriguez-Mallon A, Affuso A, Postiglione MP, Arra C, et al. An integrated regulatory network controlling survival and migration in thyroid organogenesis. Developmental biology. 2004;276(2):464-75.

66. Zhuang Y, Wu W, Liu H, Shen W. Common genetic variants on FOXE1 contributes to thyroid cancer susceptibility: evidence based on 16 studies. Tumour Biol.

2014;35(6):6159-66.

67. Swierniak M, Wojcicka A, Czetwertynska M, Dlugosinska J, Stachlewska E, Gierlikowski W, et al. Association between GWAS-Derived rs966423 Genetic Variant and Overall Mortality in Patients with Differentiated Thyroid Cancer. Clin Cancer Res. 2016;22(5):1111-9.

68. Zeng L, Sachdev P, Yan L, Chan JL, Trenkle T, McClelland M, et al. Vav3 mediates receptor protein tyrosine kinase signaling, regulates GTPase activity, modulates cell morphology, and induces cell transformation. Mol Cell Biol. 2000;20(24):9212-24.

69. Eriksson N, Tung JY, Kiefer AK, Hinds DA, Francke U, Mountain JL, et al. Novel associations for hypothyroidism include known autoimmune risk loci. PLoS One. 2012;7(4):e34442.

70. Kwak SH, Park YJ, Go MJ, Lee KE, Kim SJ, Choi HS, et al. A genome-wide association study on thyroid function and anti-thyroid peroxidase antibodies in Koreans. Hum Mol Genet. 2014;23(16):4433-42.

71. Riesco-Eizaguirre G, Santisteban P. New insights in thyroid follicular cell biology and its impact in thyroid cancer therapy. Endocr Relat Cancer. 2007;14(4):957-77.

72. Ahmed ZM, Yousaf R, Lee BC, Khan SN, Lee S, Lee K, et al. Functional null mutations of MSRB3 encoding methionine sulfoxide reductase are associated with human deafness DFNB74. Am J Hum Genet. 2011;88(1):19-29.

73. Lee BC, Dikiy A, Kim HY, Gladyshev VN. Functions and evolution of

selenoprotein methionine sulfoxide reductases. Biochim Biophys Acta. 2009;1790(11):1471-7.

74. Negro R. Selenium and thyroid autoimmunity. Biologics. 2008;2(2):265-73.

75. Panagopoulos I, Bjerkehagen B, Gorunova L, Berner JM, Boye K, Heim S. Several fusion genes identified by whole transcriptome sequencing in a spindle cell sarcoma with rearrangements of chromosome arm 12q and MDM2 amplification. International journal of oncology. 2014;45(5):1829-36.

76. Porcu E, Medici M, Pistis G, Volpato CB, Wilson SG, Cappola AR, et al. A meta-analysis of thyroid-related traits reveals novel loci and gender-specific differences in the regulation of thyroid function. PLoS Genet. 2013;9(2):e1003266.

77. Mountjoy KG, Holdaway IM, Finlay GJ. Insulin receptor regulation in cultured human tumor cells. Cancer Res. 1983;43(10):4537-42.

78. Gross GE, Boldt DH, Osborne CK. Perturbation by insulin of human breast cancer cell cycle kinetics. Cancer Res. 1984;44(8):3570-5.

79. Vella V, Pandini G, Sciacca L, Mineo R, Vigneri R, Pezzino V, et al. A novel autocrine loop involving IGF-II and the insulin receptor isoform-A stimulates growth of thyroid cancer. J Clin Endocrinol Metab. 2002;87(1):245-54.

80. Ohta M, Inoue H, Cotticelli MG, Kastury K, Baffa R, Palazzo J, et al. The FHIT gene, spanning the chromosome 3p14.2 fragile site and renal carcinoma-associated t(3;8) breakpoint, is abnormal in digestive tract cancers. Cell. 1996;84(4):587-97.

81. Joannes A, Bonnomet A, Bindels S, Polette M, Gilles C, Burlet H, et al. Fhit

regulates invasion of lung tumor cells. Oncogene. 2010;29(8):1203-13.

82. Zou M, Shi Y, Farid NR, al-Sedairy ST, Paterson MC. FHIT gene abnormalities in both benign and malignant thyroid tumours. Eur J Cancer. 1999;35(3):467-72.

83. Yin DT, Wang L, Sun J, Yin F, Yan Q, Shen RL, et al. Homozygous deletion but not mutation of exons 5 and 8 of the fragile histidine triad (FHIT) gene is associated with features of differentiated thyroid carcinoma. Annals of clinical and laboratory science. 2010;40(3):267-72.

84. Yin DT, Wang L, Sun J, Yin F, Yan Q, Shen R, et al. Association of the promoter methylation and protein expression of Fragile Histidine Triad (FHIT) gene with the progression of differentiated thyroid carcinoma. International journal of clinical and experimental pathology. 2010;3(5):482-91.

85. Hanai N, Nagata K, Kawajiri A, Shiromizu T, Saitoh N, Hasegawa Y, et al. Biochemical and cell biological characterization of a mammalian septin, Sept11. FEBS Lett. 2004;568(1-3):83-8.

86. Kojima K, Sakai I, Hasegawa A, Niiya H, Azuma T, Matsuo Y, et al. FLJ10849, a septin family gene, fuses MLL in a novel leukemia cell line CNLBC1 derived from chronic neutrophilic leukemia in transformation with t(4;11)(q21;q23). Leukemia. 2004;18(5):998-1005.

87. Kim NG, Rhee H, Li LS, Kim H, Lee JS, Kim JH, et al. Identification of MARCKS, FLJ11383 and TAF1B as putative novel target genes in colorectal carcinomas with microsatellite instability. Oncogene. 2002;21(33):5081-7.

88. Vaidya B, Kendall-Taylor P, Pearce SH. The genetics of autoimmune thyroid disease. J Clin Endocrinol Metab. 2002;87(12):5385-97.

89. Czene K, Lichtenstein P, Hemminki K. Environmental and heritable causes of cancer among 9.6 million individuals in the Swedish Family-Cancer Database. Int J Cancer. 2002;99(2):260-6.

90. Grimm C, Kraft R, Sauerbruch S, Schultz G, Harteneck C. Molecular and functional characterization of the melastatin-related cation channel TRPM3. J Biol Chem. 2003;278(24):21493-501.

91. Wagner TF, Loch S, Lambert S, Straub I, Mannebach S, Mathar I, et al. Transient receptor potential M3 channels are ionotropic steroid receptors in pancreatic beta cells. Nat Cell Biol. 2008;10(12):1421-30.

92. Hall DP, Cost NG, Hegde S, Kellner E, Mikhaylova O, Stratton Y, et al. TRPM3 and miR-204 establish a regulatory circuit that controls oncogenic autophagy in clear cell renal cell carcinoma. Cancer Cell. 2014;26(5):738-53.

93. Son HY, Hwangbo Y, Yoo SK, Im SW, Yang SD, Kwak SJ, et al. Genome-wide association and expression quantitative trait loci studies identify multiple susceptibility loci for thyroid cancer. Nat Commun. 2017;8:15966.

94. Wang J, Song J, An C, Dong W, Zhang J, Yin C, et al. A 130-kDa protein 4.1B regulates cell adhesion, spreading, and migration of mouse embryo fibroblasts by influencing actin cytoskeleton organization. J Biol Chem. 2014;289(9):5925-37.

95. Parra M, Gascard P, Walensky LD, Gimm JA, Blackshaw S, Chan N, et al. Molecular and functional characterization of protein 4.1B, a novel member of the

protein 4.1 family with high level, focal expression in brain. J Biol Chem. 2000;275(5):3247-55.

96. Wang Z, Zhang J, Ye M, Zhu M, Zhang B, Roy M, et al. Tumor suppressor role of protein 4.1B/DAL-1. Cell Mol Life Sci. 2014;71(24):4815-30.

97. Gorski M, van der Most PJ, Teumer A, Chu AY, Li M, Mijatovic V, et al. 1000 Genomes-based meta-analysis identifies 10 novel loci for kidney function. Sci Rep. 2017;7:45040.

98. Gottlieb DJ, Hek K, Chen TH, Watson NF, Eiriksdottir G, Byrne EM, et al. Novel loci associated with usual sleep duration: the CHARGE Consortium Genome-Wide Association Study. Mol Psychiatry. 2015;20(10):1232-9.

99. Teumer A, Rawal R, Homuth G, Ernst F, Heier M, Evert M, et al. Genome-wide association study identifies four genetic loci associated with thyroid volume and goiter risk. Am J Hum Genet. 2011;88(5):664-73.

100. Zhan M, Chen G, Pan CM, Gu ZH, Zhao SX, Liu W, et al. Genome-wide association study identifies a novel susceptibility gene for serum TSH levels in Chinese populations. Hum Mol Genet. 2014;23(20):5505-17.

101. Taylor PN, Porcu E, Chew S, Campbell PJ, Traglia M, Brown SJ, et al. Whole-genome sequence-based analysis of thyroid function. Nat Commun. 2015;6:5681.

102. Horvath E, Majlis S, Rossi R, Franco C, Niedmann JP, Castro A, et al. An ultrasonogram reporting system for thyroid nodules stratifying cancer risk for clinical management. J Clin Endocrinol Metab. 2009;94(5):1748-51.

103. Gharib H, Papini E, Paschke R, Duick DS, Valcavi R, Hegedus L, et al. American Association of Clinical Endocrinologists, Associazione Medici Endocrinologi, and EuropeanThyroid Association Medical Guidelines for Clinical Practice for the Diagnosis and Management of Thyroid Nodules. Endocr Pract. 2010;16 Suppl 1:1-43.

104. Yi KH. The Revised 2016 Korean Thyroid Association Guidelines for Thyroid Nodules and Cancers: Differences from the 2015 American Thyroid Association Guidelines. Endocrinol Metab (Seoul). 2016;31(3):373-8.

105. Cibas ES, Ali SZ, Conference NCITFSotS. The Bethesda System For Reporting Thyroid Cytopathology. Am J Clin Pathol. 2009;132(5):658-65.

106. Durante C, Costante G, Lucisano G, Bruno R, Meringolo D, Paciaroni A, et al. The natural history of benign thyroid nodules. JAMA. 2015;313(9):926-35.

107. Lee S, Skelton TS, Zheng F, Schwartz KA, Perrier ND, Lee JE, et al. The biopsy-proven benign thyroid nodule: is long-term follow-up necessary? J Am Coll Surg. 2013;217(1):81-8; discussion 8-9.

국문초록

갑상선암과 결절에 대한 전장 유전체 연관 및 발현 양적 형질 유전자좌 연구

서울대학교 대학원 의학과 중개의학 전공

황 보 율

갑상선암은 한국에서 가장 흔한 내분비암이며 갑상선 결절은 가장 흔한 내분비 질환이다. 두가지 질환 모두 높은 유전성을 보인다. 몇몇의 갑상선암에 대한 전장 유전체 연관 연구가 서양인들에게서 이루어졌고, 분화갑상선암에 대한 감수성 유전자좌를 발굴하였다. 그러나 아시아인에 대한 전장 유전체

연관 연구는 수행된 바 없으며, 갑상선 결절에 대한 유전적 연구는 없었으며 이와 관련된 유전자와 갑상선암과의 관련성도 여전히 알 수 없는 상태이다.

따라서, 본 연구에서는 1,085 명의 분화 갑상선암과 8,884 명의 대조군으로 전장 유전체 연관 분석 및 재현 연구를 수행하였고, 그 결과를 발현 양적 형질 유전자좌 연구 및 임상 발현형질을 통해서 검증하였다. 가장 뚜렷한 관련성은 보이는 유전자좌는 *NRG1* 유전자였으며 (rs6996585, P=1.08×10^{-10}), 이 SNP 은 *NRG1* 의 발현과도 관련성이 있었다. 부가적으로 이전에 보고되었던 유전자좌 (*FOXE1, NKX2-1, DIRC3*)를 확인하였으며 7 개의 유전자좌 (*VAV3, PCNXL2, INSR, MRSB3, FHIT, SEPT11, SLC24A6*)를 새롭게 발견하였다. 또한, 분화갑상선암과 관련된 유전변이가 암의 종류 및 인종에 따라서 다른 영향을 가지는 것을 확인하였다.

또한 갑상선결절에 대한 3 단계의 전장 유전체 연관 분석을 시행하였다. 발견 단계의 전장 유전체 스캔을 인구 기반 코호트의 811 명의 갑상선 결절군과 691 명의 정상 갑상선군에서

수행되었다. 재현 연구는 건강검진 대상자에서 총 1981 명의 결절군과 3100 명의 정상군에서 수행되었으며 발현 양적 형질 유전자좌 분석도 공공데이터를 통해서 수행되었다. 가장 유의한 관련성은 결합분석 (OR=1.26, $P = 6.12 \times 10^{-8}$) 및 메타분석 (OR = 1.28, $P = 2.11 \times 10^{-8}$) 결과 *TRPM3* (rs4745021) 유전자에서 관찰되었다. *MBIP/NKX2-1* 변이는 재현이 되었으나 전장 유전체 유의성을 보이지 못했다. 발현 양적 형질 유전자좌 분석에서 *TRPM3* 의 발현은 갑상선조직에서 rs4745021 유전자형과 관련성이 있었다.

분화갑상선암에 대한 전장 유전체 연관 분석 결과는 갑상선암의 발생에서 유전적 기여에 대한 이해할 수 있게 해주었으며, 갑상선 결절에 대한 전장 유전체 연구를 통해 갑상선 결절은 갑상선암과 차별되는 유전적 특징을 가지고 있음을 확인하였다.

· 본 박사학위논문은 출판된 다음의 연구 논문을 기반으로 작성되었습니다.

· Genome-Wide Association and Expression Quantitative Trait Loci Studies Identify Multiple Susceptibility Loci for Thyroid Cancer. Nat Commun. 2017 Jul 13;8:15966.

· Genome-Wide Association Study Reveals Distinct Genetic Susceptibility of Thyroid Nodules

From Thyroid Cancer. J Clin Endocrinol Metab. 2018 Dec 1;103(12):4384-4394.

· Genome-Wide Association Studies of Autoimmune Thyroid Diseases, Thyroid Function, and Thyroid Cancer. Endocrinol Metab (Seoul). 2018 Jun; 33(2): 175–184.

주요어: 갑상선암, 갑상선결절, 전장 유전체 연관 분석, 발현 양적 형질 유전자좌, 유전적 감수성

학번: 2015-30903